Competitive Selection of Ephemeral Relays in Wireless Networks

K. P. Naveen, Eitan Altman, *Fellow, IEEE,* and Anurag Kumar, *Fellow, IEEE*

***Abstract*—We consider a setting in which two nodes (referred to as forwarders) compete to choose a relay node from a set of relays, as they ephemerally become available (e.g., wake up from a sleep state). Each relay, when it arrives, offers a (possibly different) "reward" to each forwarder. Each forwarder's objective is to minimize a combination of the delay incurred in choosing a relay and the reward offered by the chosen relay. As an example, we develop the reward structure for the specific problem of geographical forwarding over a network of sleep-wake cycling relays.**

We study two variants of the generic relay selection problem, namely, the completely observable (CO) case where, when a relay arrives, both forwarders get to observe both rewards, and the partially observable (PO) case where each forwarder can only observe its own reward. Formulating the problem as a two person stochastic game, we characterize solution in terms of Nash Equilibrium Policy Pairs (NEPPs). For the CO case we provide a general structure of the NEPPs. For the PO case we prove that there exists an NEPP within the class of threshold policy pairs.

We then consider the particular application of geographical forwarding of packets in a shared network of sleep-wake cycling wireless relays. For this problem, for a particular reward structure, using realistic parameter values corresponding to TelosB wireless mote, we numerically compare the performance (in terms of cost to both forwarders) of the various NEPPs and draw the following key insight: even for moderate separation between the two forwarders, the performance of the various NEPPs is close to the performance of a simple strategy where each forwarder behaves as if the other forwarder is not present. We also conduct simulation experiments to study the end-to-end performance of the simple forwarding policy.

***Index Terms*—Competitive relay selection, geographical forwarding, stochastic games, Bayesian games.**

I. Introduction

We are concerned in this paper with a class of resource allocation problems in wireless networks, in which competing nodes need to acquire a resource, such as a physical radio relay (see the geographical forwarding example later in Section III) or a channel (as in a cognitive radio network [1], [2]), when a sequence of such resources "arrive" over time, and are available only fleetingly for acquisition. In this paper, formulating such a problem for two nodes as a stochastic game, we consider the completely observable and partially observable cases, and provide characterizations of the Nash Equilibrium Policy Pairs (NEPP). We provide numerical results, and insights therefrom, for a specific reward structure derived from the problem of geographical forwarding in sleep-wake cycling networks.

The Geographical Forwarding Context: With the increasing importance of "smart" utilization of our limited resources (e.g., energy and clean water) there is a need for instrumenting our buildings and campuses with wireless sensor networks. As awareness grows and sensing technologies emerge, new applications will be implemented. While each application will require different sensors and back-end analytics, the availability of a common wireless network infrastructure will promote the quick deployment of new applications. One approach for building such an infrastructure, say, in a large building setting, would be to deploy a large number of relay nodes, and employ the idea of geographical forwarding. If the phenomena to be monitored are slowly varying over time, the traffic on the network can be assumed to be light. In addition, such applications are *delay tolerant,* thus accommodating the approach of *opportunistic* geographical forwarding over *sleep-wake cycling* networks [3], [4].

Sleep-wake cycling is an approach whereby, to conserve the relay battery power, their radios are kept turned OFF, while coming ON periodically to provide opportunities for packet forwarding. The problem of forwarding in such a setting was explored in [3], [4], where the formulation was limited to a *single* alarm packet flowing through the network. Whereas the emphasis in [3] was to develop an end-to-end optimal forwarding algorithm, thus requiring a global organization step, in [4], which is our prior work on this problem, we sought a locally-optimal forwarding heuristic. End-to-end forwarding was achieved by applying the local heuristic at each forwarding step. We found that, over certain range of operation, the performance obtained by the heuristic is comparable with the optimal solution provided by [3].

In the setting discussed above, even though the traffic is light, there is still a chance that there is more than one forwarder seeking a relay from among a set of potential relays. There then arises the problem of assigning the relays, as they wake-up, to one or the other of the forwarders. This, thus, is an extension of the local forwarding problem discussed in [4]. Formally, the local forwarding problem we consider in this paper is the following. There are two forwarders each of

Author's addresses: K. P. Naveen, INRIA Saclay, Palaiseau 91120, France (email: naveenkp@inria.fr); Anurag Kumar, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (email: anurag@ece.iisc.ernet.in); Eitan Altman, INRIA, Sophia Antipolis 06902, France (email: Eitan.Altman@inria.fr).

This work was done when the first author was a Ph.D student at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India.

This work was supported by the following projects: Indo-French Center for the Promotion of Advanced Research under Project 4000IT1, and the IFCAM (Indo-French Centre for Applied Mathematics) program; Department of Science and Technology (DST) through project WINSON (Wireless Networks and Techniques with Applications to Social Needs) and a J.C. Bose National Fellowship; and the INRIA Associates Project GANESH (Games Optimization and Analysis of Networks: Theory and Applications).

which has to choose a relay node to forward its packet to. The relays are waking up sequentially over time. Whenever a relay wakes up, each forwarder first evaluates the relay based on a reward metric (which could be a function of the progress, towards the sink, made by the relay, and the power required to get the packet to the relay [4]), and then decides whether to compete (with the other forwarder) for this relay or continue to wait for further relays to wake-up. Such a geographical forwarding setting will serve as an example application of the stochastic game formulation developed in this paper.

Outline and Our Contributions: We will describe a general system model in Section II, following which, in Section III, we will discuss a geographical forwarding problem as an example. Related work will be presented in Section IV. In Sections V and VI we will study two variants of the problem (of progressive complexity), namely, one where *complete information* is available to both forwarders and one with only *partial information*. We will use stochastic game theory to obtain solution in terms of (stationary) Nash Equilibrium Policy Pairs (NEPPs). We will briefly study a cooperative setting in Section VII, and obtain the Pareto optimal performance curve which provides a benchmark for the NEPPs. The following are our main technical contributions:

- For the problem with complete information we obtain results illustrating the structure of NEPPs (Theorem 2)
- For the partial information case we prove the existence of a NE strategy (for a certain Bayesian game) within the class of threshold strategies (Theorem 4). This result will enable us to construct NEPPs for this case.
- In Section VIII we provide a simulation study of the use of our formulation in the context of geographical forwarding. Using realistic parameters from the popular TelosB wireless mote, we make the following interesting observation: even for moderate separation between the two forwarders, the performance of all the NEPPs is close to the performance of a *simple strategy* where each forwarder behaves as if it is alone.

We will finally draw our conclusions in Section IX. For the ease of presentation we have moved most of the proofs to the Appendix.

II. System Model

Let $\mathscr{F}_1$ and $\mathscr{F}_2$ denote the two competing nodes (i.e., players in game theoretic terms), referred to as the *forwarders*. We will assume that there are an infinite number of *relay nodes* (or *resources* in general) that are arriving sequentially at times $\{W_k : k \geq 0\}$, which are the points of a Poisson process of rate $\frac{1}{\tau}$. Thus, the *inter-"arrival"* times between successive relays, $U_k := W_k - W_{k-1}$, are i.i.d. (independent and identically distributed) exponential random variables of mean τ. We will refer to the relay that arrives at the instant W_k as the *k-th relay*. Further, the k-th relay is only ephemerally available at the instant W_k.

When a relay arrives, either of the forwarders can compete for it, thereby obtaining a *reward*. Let $R_{\rho,k}$, $\rho = 1, 2$, denote the *reward* offered by the k-th relay to $\mathscr{F}_\rho$ (an example reward structure will be discussed in Section III). The rewards $R_{\rho,k}$ ($\rho = 1, 2;\ k \geq 1$) can take values from a finite set $\mathcal{R} = \{r_1, r_2, \cdots, r_n\}$, where $r_1 = -\infty$ and $r_i < r_j$ for $i < j$. The reward pairs $(R_{1,k}, R_{2,k})$ are i.i.d. across k, with their common joint p.m.f. (probability mass function) being $p_{R_1,R_2}(\cdot,\cdot)$, For notational simplicity we will denote $p_{R_1,R_2}(r_i, r_j)$ as simply $p_{i,j}$. Further, let $p_i^{(1)}$ and $p_j^{(2)}$ denote the marginal p.m.f.s of $R_{1,k}$ and $R_{2,k}$, respectively. Thus, $p_i^{(1)} = \sum_{j=1}^n p_{i,j}$ and $p_j^{(2)} = \sum_{i=1}^n p_{i,j}$.

Actions and Consequences: First we will study (in Section V) a *completely observable case* where the reward pair $(R_{1,k}, R_{2,k})$ is revealed to both the forwarders. Later, in Section VI, we will consider a more involved (albeit more practical) *partially observable case* where only $R_{1,k}$ is revealed to $\mathscr{F}_1$, and $R_{2,k}$ is revealed to $\mathscr{F}_2$. However in either case, each time a relay arrives, the two forwarders have to independently choose between one of the following actions:

- s: *stop* and forward the packet to the current relay, or
- c: *continue* to wait for further relays to arrive.

Suppose both forwarders choose to stop, then with probability (w.p.) ν_1, $\mathscr{F}_1$ gets the relay in which case $\mathscr{F}_2$ has to continue alone, while with the remaining probability ($\nu_2 = 1-\nu_1$) $\mathscr{F}_2$ gets the relay and $\mathscr{F}_1$ continues alone. ν_ρ ($\rho = 1, 2$) could be thought of as the probability that $\mathscr{F}_\rho$ will win the contention when both forwarders attempt simultaneously. For mathematical tractability we will assume that the forwarders make their decision instantaneously at the relay arrival instants. Further, if a relay is not chosen by either forwarder (i.e., if both forwarders choose to continue) we will assume that the relay disappears and is not available for further use.

System State and Forwarding Policy: For the CO case, $(R_{1,k}, R_{2,k})$ can be regarded as the *state* of the system at stage k, provided both forwarders have not *terminated* (i.e., chosen a relay) yet. When one of the forwarder, say $\mathscr{F}_2$, terminates, we will represent the system state as $(R_{1,k}, \boldsymbol{t})$. Similarly, let $(\boldsymbol{t}, R_{2,k})$ and $(\boldsymbol{t}, \boldsymbol{t})$ represents the state of the system when only $\mathscr{F}_1$ has terminated and when both forwarders have terminated, respectively. Formally, we can define the *state space* to be

$$\mathcal{X} = \Big\{(r_i, r_j), (r_i, \boldsymbol{t}), (\boldsymbol{t}, r_j), (\boldsymbol{t}, \boldsymbol{t}) : r_i, r_j \in \mathcal{R}\Big\}. \tag{1}$$

Given a discrete set $\mathcal{S}$, let $\Delta(\mathcal{S})$ denote the set of all p.m.f.s on $\mathcal{S}$. We now have the following definition.

Definition 1: A *forwarding policy* π is a mapping, $\pi : \mathcal{X} \to \Delta(\{\mathsf{s},\mathsf{c}\})$, such that $\mathscr{F}_1$ (or $\mathscr{F}_2$) using π will choose action s or c according to the p.m.f. $\pi(x_k)$ when the state of the system at stage $k \geq 1$ is $x_k \in \mathcal{X}$. A *policy pair* (π_1, π_2) is a tuple of policies such that $\mathscr{F}_1$ uses π_1 and $\mathscr{F}_2$ uses π_2.

Note that we have restricted to the class of stationary policies only. We will denote this class of policies as Π_S.

Problem Formulation: Suppose the forwarders use a policy pair (π_1, π_2), and let $x \in \mathcal{X}$ be the state of the system at stage 1. Let K_ρ, $\rho = 1, 2$, denote the (random) stage at which $\mathscr{F}_\rho$ forwards its packet. Then, the delay incurred by $\mathscr{F}_\rho$ ($\rho = 1, 2$), starting from the instant $W_1 = U_1$ (first relay's arrival instant), is $D_{K_\rho} = U_2 + \cdots + U_{K_\rho}$, and the reward accrued is R_{ρ,K_ρ}. Let $\mathbb{E}^x_{\pi_1,\pi_2}[\cdot]$ denote the expectation operator corresponding to the probability law, $\mathbb{P}^x_{\pi_1,\pi_2}$, governing the system dynamics

when the policy pair used is (π_1, π_2) and the initial state is x. Then, the expected total cost incurred by $\mathscr{F}_\rho$ is

$$J^{(\rho)}_{\pi_1,\pi_2}(x) = \mathbb{E}^x_{\pi_1,\pi_2}\Big[D_{K_\rho} - \eta_\rho R_{\rho,K_\rho}\Big], \tag{2}$$

where $\eta_\rho > 0$ is the multiplier used to trade-off between delay and reward.

Definition 2: We say that a policy pair (π_1^*, π_2^*) is a Nash equilibrium policy pair (NEPP) if, for all $x \in \mathcal{X}$, $J^{(1)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(1)}_{\pi_1,\pi_2^*}(x)$ for any policy $\pi_1 \in \Pi_S$, and $J^{(2)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(2)}_{\pi_1^*,\pi_2}(x)$ for any policy $\pi_2 \in \Pi_S$. Thus, a unilateral deviation from an NEPP is neither beneficial for $\mathscr{F}_1$ nor for $\mathscr{F}_2$.

Our objective will be to characterize the solution in terms of NEPPs.

III. Geographical Forwarding Example

Before proceeding further, in this section, as a motivating example, we will construct a reward structure corresponding to the problem of *geographical forwarding*[1] in sleep-wake cycling wireless networks. Let $\mathscr{F}_1$ and $\mathscr{F}_2$ actually represent two forwarding nodes in a wireless network. As shown in Fig. 1, let v_1 and v_2 denote their respective locations. A sink node is located at v_0. Let d denote the range of both the forwarders. Given any location $\ell \in \Re^2$, we define the *progress*, $Z_\rho(\ell)$, made by location ℓ with respect to (w.r.t.) $\mathscr{F}_\rho$ as

$$Z_\rho(\ell) \quad = \quad \| v_\rho - v_0 \| - \| \ell - v_0 \| \tag{3}$$

where $\| \cdot \|$ denotes the Euclidean norm. Thus, $Z_\rho(\ell)$ is simply the difference between $\mathscr{F}_\rho$-to-sink and ℓ-to-sink distances. A positive value of $Z_\rho(\ell)$ implies that location ℓ is closer to the sink than $\mathscr{F}_\rho$. Now, define the forwarding region, $\mathcal{L}_\rho$, of $\mathscr{F}_\rho$ as the set of all locations that lie within the range of $\mathscr{F}_\rho$ and make non-negative progress w.r.t. $\mathscr{F}_\rho$, i.e., denoting $D_\rho(\ell) = \| \ell - v_\rho \|$ to be the distance between ℓ and $\mathscr{F}_\rho$,

$$\mathcal{L}_\rho \quad = \quad \Big\{\ell : D_\rho(\ell) \leq d, Z_\rho(\ell) \geq 0\Big\}. \tag{4}$$

Let $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ denote the combined forwarding region of the two forwarders. As depicted in Fig. 1, we will discretize $\mathcal{L}$ into a grid of finite set of m locations $\{\ell_1, \ell_2, \cdots, \ell_m\}$. Thus, from here on, whenever we refer to a location ℓ we mean it to be one of the above m locations.

Sleep-Wake Process: Without loss of generality, we will assume that at time 0 each forwarder is holding an alarm packet which has to forwarded to a downstream *relay node* (i.e., a node in its forwarding region). Since the relays are sleep-wake cycling, each forwarder has to wait until a "good" relay wakes up (the goodness of a relay will be based on the reward metric to be discussed in this section).

A practical approach for sleep-wake cycling is the *asynchronous periodic sleep-wake process* [3], [4], where each relay i wakes up at the periodic instants $\{T_i + kT : k \geq 0\}$ with $\{T_i\}$ being i.i.d. (independent and identically distributed) uniform on $[0, T]$ (T is referred to as the sleep-wake cycling period). Now, for dense networks where N is large, if T scales

[1]Geographical forwarding [5], [6], also known as location based routing, is a forwarding technique where the assumption is that each node knows its location as well as the location of the sink node.

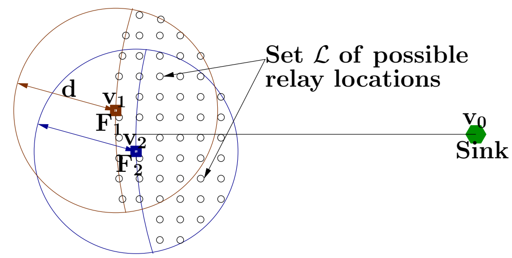


Fig. 1. One-hop forwarding scenario: v_0, v_1, and v_2 are the locations of the sink, $\mathscr{F}_1$, and $\mathscr{F}_2$, respectively; d is the range of each forwarder. Possible relay locations are shown as $\circ$.

with N such that $\frac{N}{T} = \frac{1}{\tau}$ as $N \to \infty$, then the aggregate point process of relay wake-up instants converges to a Poisson process of rate $\frac{1}{\tau}$ [7]. This observation motivates us to model the *aggregate* point process of wake-up instants of relays as a Poisson point process. Furthermore, the Poisson point process assumption renders our problem analytically tractable, leading to interesting structural results.

Thus, formally, we model the sleep-wake cycling by assuming that there are an infinite number of relays waking up (within the combined forwarding region $\mathcal{L}$) sequentially at the times $\{W_k : k \geq 0\}$ which are the points of a Poisson process of rate $\frac{1}{\tau}$ (thus, a new relay wakes up at each instant W_k). Let $L_k \in \mathcal{L}$ denote the location of the k-th relay (i.e., the relay waking up at the instant W_k). The locations $\{L_k : k \geq 1\}$ are i.i.d. random variables with their common p.m.f. (probability mass function) being q, i.e., $\mathbb{P}(L_k = \ell) = q_\ell$.

Channel Model: We will use the following standard model to obtain the transmission power required by $\mathscr{F}_\rho$ to achieve an SNR (signal to noise ratio) constraint of Γ at the k-th relay:

$$P_{\rho,k} = \frac{\Gamma N_0}{G_{\rho,k}}\left(\frac{D_\rho(L_k)}{d_{ref}}\right)^\xi \tag{5}$$

where, N_0 is the receiver noise variance, $D_\rho(L_k)$ is the distance between $\mathscr{F}_\rho$ and the k-th relay whose location is L_k, $G_{\rho,k}$ is the gain of the channel between $\mathscr{F}_\rho$ and the k-th relay, ξ is the path-loss attenuation factor, and d_{ref} is the far-field reference distance beyond which the above expression is valid [8], [9] (our discretization of $\mathcal{L}$ is such that the distance between $\mathscr{F}_\rho$ and any $\ell \in \mathcal{L}$ is more than d_{ref}).

We will assume that the set of channel gains $\{G_{\rho,k} : k \geq 1, \rho = 1, 2\}$ are i.i.d. taking values from a finite set $\mathcal{G}$. Also, let P_{max} denote the maximum transmit power with which the two forwarders can transmit, i.e., if $P_{\rho,k} > P_{max}$ then $\mathscr{F}_\rho$ cannot forward its packet to the k-th relay. Further, we assume that the range d (recall Fig. 1) is such that if the k-th relay is outside the range of $\mathscr{F}_\rho$ (i.e., $D_\rho(L_k) > d$), then for any $G_{\rho,k} \in \mathcal{G}$, $P_{\rho,k} > P_{max}$, so that $\mathscr{F}_\rho$ cannot forward to a relay outside its range. Transmitting to a relay inside its range is possible, however, provided the channel gain is good enough so that the power required is less than P_{max}.

Relay Rewards: Finally, combining progress and power, we will define the reward offered by the k-th relay to $\mathscr{F}_\rho$ as,

$$R_{\rho,k} \quad = \quad \begin{cases} \frac{Z_\rho(L_k)^a}{P_{\rho,k}^{(1-a)}} & \text{if } P_{\rho,k} \leq P_{max} \\ -\infty & \text{otherwise,} \end{cases} \tag{6}$$

where $a \in [0, 1]$ is the parameter used to trade-off between progress and power. The reward being inversely proportional to power is clear because it is advantageous to use low power to get the packet across; $R_{\rho,k}$ is made proportional to $Z_\rho(L_k)$ to promote progress towards the sink while choosing a relay for the next hop.

We will use the above reward structure for conducting numerical and simulation experiments in Section VIII. However, it is important to note that all our analysis in the subsequent sections hold for the general model introduced in Section II.

IV. Related Work

We will first make an important comparison with our prior work on the topic, before proceeding to discuss general literature from the area of geographical forwarding in wireless networks. Our problem can also be considered as a variant of the *asset selling problem* studied in the operations research literature; we will discuss related work from this field as well. Finally, we survey literature from the area of stochastic games.

Our Prior Work: Problem of relay selection, but by a single forwarder (i.e., the non-competitive version), has been extensively studied by us, starting from a simple model where the number of relays is exactly known to the forwarder to the one where only a belief is known [4]. We have also studied a variant with channel probing where the relay rewards are not immediately revealed to the forwarder; instead the forwarder can choose to learn the reward values by paying an additional cost [10].

The basic version of our model [4, Section 6] comprises only one forwarder and a finite number of relay N; however, in the basic model we allow for the forwarder to recall a previous relay unlike here where recalling is not allowed. For this basic model, the solution is completely in terms of a single threshold α: forward to the first relay whose reward is more than α; at the last stage choose the best relay irrespective of its reward value. From [4, Section 6], we further know that the value of α does not dependent on N, and hence the solution to the version of the basic model with infinite number of relays, should still be same. Furthermore, in the infinite horizon model there is no advantage in recalling the best relay since there is no last stage. Thus, one can argue that the solution to the *infinite horizon basic relay selection model, without recall,* should also be characterized by the same threshold α. Here, we will formally show that this is in fact the solution for one forwarder when the other forwarder has already terminated (Lemma 1). However when both the forwarders are present, the solution is more involved (studied in Section V-B). Thus, the competitive model studied here is a generalization of the aforementioned version of the basic relay selection model.

Geographical Forwarding: The problem of choosing a next-hop relay arises in the context of *geographical forwarding*; *geographical forwarding* [5], [6] is a forwarding technique where the prerequisite is that the nodes know their respective locations as well as the sink's location. The method of geographical forwarding was already envisioned in the 80's in the context of routing in packet radio networks (PRNs) [11], [12]. One of the simplest geographical forwarding technique is the greedy algorithm where each node forwards to a neighbor in its communication region which makes maximum progress towards the sink. This greedy algorithm is referred to as the *MFR (Max Forward within Radius)* routing in [11]. Akin to MFR is the *NFP (Nearest with Forward Progress)* proposed in [12] where a node with a positive progress, and closest to the transmitting node is chosen. A generalization of MFR and NFP routing is to randomly choose any neighbor which makes a positive progress towards the sink [13].

More recently, there are work that apply geographical forwarding for routing in sleep-wake cycling networks. For instance, Zorzi and Rao in [14] propose an algorithm called GeRaF (Geographical Random Forwarding) which, at each forwarding stage, chooses the relay making the largest progress. For a sleep-wake cycling network, Liu et al. in [15] propose a relay selection approach as a part of CMAC, a protocol for geographical packet forwarding. Under CMAC, node i chooses an r_0 that minimizes the expected normalized latency (which is the average ratio of one-hop delay and progress). Akin to the relay selection problem is the problem of channel selection [16], [17] where a transmitter, given several channels, has to choose one for its transmissions. Analogous to rewards in our case, the transmitter's decision is based on the throughput the transmitter can achieve on a channel. Links to more literature on similar work from the context of wireless networks can be found in [4]. However all these work do not consider the competitive scenario like ours.

Asset Selling Problem: Finally, our relay selection problem can be considered to be equivalent to the *asset selling problem*, which is a class of the optimal stopping problems studied in the operations research literature (other examples of stopping problems include the *secretary problem* [18], *bandit problem* [19], etc). The basic asset selling problem [20, Section 4.4] [21] comprises a single seller (analogous to a forwarder in our model) and a sequence of i.i.d. offers (rewards in our case). The seller's objective is to choose an offer so as to maximize a combination of the offer value and the number of previous offers rejected. Over the years, several variants of the basic problem have been studied. For instance, In [22], David and Levi consider a model in which the offers arrive at the points of a renewal process. Kang in [23] has considered a model where a cost has to be paid to recall the previous best offer; see [23] for further references to literature on models with uncertain recall. Variants with unknown offer (or reward) distribution, or one where a parameter of the offer distribution is unknown have been studied in [24], [25].

Our competitive model here can be considered as a game variant of the basic asset selling problem, where the two forwarders are analogous to the sellers and the reward values are analogous to the offers. Although one game variant has been studied by Nakagami in [26], the specific cost structure in our problem enables us to prove results such as the existence of Nash equilibrium policy pair within the class of threshold rules (Theorem 4). Further, we also study a completely observable case which is not considered in [26].

Similarly, literature is available on the game version of the secretary problem [27], [28], but these consider the simpler case where the reward offered by an arriving secretary (or

resource) to both players is the same. Moreover, the objective in the secretary problem is to maximize the probability of choosing the best secretary (resource), which is in contrast to our setting (asset selling) which involves a trade-off between selection delay and reward. Further, a partially observable scenario is not studied in these work.

Stochastic Games: Stochastic games can be considered as a generalization of Markov decision processes (MDPs), in the sense that a stochastic game comprises multiple agents (in contrast to a single agent in an MDP), who jointly control the state of the system while individually incurring a cost in doing so. Several references [29]–[34] are available on the topic starting from the seminal work by Shapley [35]. However, most of these work study either discounted or average cost objectives, unlike our problem which falls within the realm of total-cost transient stochastic games (or stopping games [36, Part III]). Our formulation can be alternatively thought of as a quitting game [37]. However, we have introduced state transitions and state dependent quitting cost which are not considered in the model studied in [37].

In summary, to the best of our knowledge, the model proposed in this paper along with the structural results we have derived, are new contributions to the field of stopping games.

V. Completely Observable (CO) Case

For the CO model we assume that the reward pair, $(R_{1,k}, R_{2,k})$, of the k-th relay is entirely revealed to both the forwarders. Recalling the geographical forwarding example in Section III, this case would model the scenario where the reward is simply the progress, $Z_\rho(L_k)$, the relay makes towards the sink, i.e., if $a = 1$ in (6). Thus, observing the location L_k of the k-th relay, both forwarders (assuming that both a-priori know the locations v_1, v_2 and v_0; see the following remark) can entirely compute $(R_{1,k}, R_{2,k})$.

Remark: Justification for knowing the locations is as follows. All the nodes are equipped with GPS (Global Positioning System) devices, using which each node can know its own location. Next, the sink being a fixed node, its location is already made available to all the nodes before deployment. Finally, each forwarder's knowledge of the other's location can be acquired when both forwarders broadcast control packets in response to the control packet transmitted by the first relay.

We will now proceed to formulate the completely observable case as a stochastic game. Using a key theorem from the book by Filar and Vrieze on *Competitive Markov Decision Processes* [29], we will characterize the structure of NEPPs.

A. Stochastic Game Formulation

Limiting ourselves to the case of finite set of states and finite action sets, formally a stochastic game can be represented by a tuple $(\mathcal{N}, \mathcal{X}, \{\mathcal{A}_\rho\}, T, \{g_\rho\})$ where,

- $\mathcal{N}$ is the set of agents or players,
- $\mathcal{X}$ is the finite set of system states,
- $\mathcal{A} = \times_{\rho\in\mathcal{N}} \mathcal{A}_\rho$ is the joint-action space with A_ρ representing the finite action set of agent ρ,
- $T : \mathcal{X} \times \mathcal{A} \to \Delta(\mathcal{X})$ (the set of all p.m.f.s on $\mathcal{X}$) is the probability transition kernel, i.e., $T(x'|x,a)$ is the probability that the next state is x' given that the current state is x and the current joint-action is $a = (a_\rho : \rho \in \mathcal{N})$,
- $g_\rho : \mathcal{X} \times \mathcal{A} \to \Re$ is the (expected) one-step-cost function of agent ρ.

We will now identify each of these components for our problem. The two forwarders, $\mathscr{F}_1$ and $\mathscr{F}_2$, are the players (i.e., $\mathcal{N} = \{F_1, F_2\}$), and $\mathcal{X}$ in (1) is the state space. The action sets are $\mathcal{A}_1 = \mathcal{A}_2 = \{\mathtt{s},\mathtt{c}\}$.

Transition Probabilities: Recall that $p_{i,j}$ is the joint p.m.f of $(R_{1,k}, R_{2,k})$, $p_i^{(1)}$ and $p_j^{(2)}$ are the marginal p.m.f.s of $R_{1,k}$ and $R_{2,k}$, respectively, and ν_ρ ($\rho = 1,2$) is the probability that $\mathscr{F}_\rho$ will win the contention if both forwarders cooperate. Now, the transition probability when the current state is of the form $x = (r_i, r_j)$ can be written as,

$$T(x'|x,a) = \begin{cases} p_{i',j'} & \text{if } a = (\mathtt{c},\mathtt{c}), x' = (r_{i'}, r_{j'}) \\ p_{i'}^{(1)} & \text{if } a = (\mathtt{c},\mathtt{s}), x' = (r_{i'}, \boldsymbol{t}) \\ p_{j'}^{(2)} & \text{if } a = (\mathtt{s},\mathtt{c}), x' = (\boldsymbol{t}, r_{j'}) \\ \nu_2 p_{i'}^{(1)} & \text{if } a = (\mathtt{s},\mathtt{s}), x' = (r_{i'}, \boldsymbol{t}) \\ \nu_1 p_{j'}^{(2)} & \text{if } a = (\mathtt{s},\mathtt{s}), x' = (\boldsymbol{t}, r_{j'}) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Note that when the joint-action is $(\mathtt{s},\mathtt{s})$, $\nu_2 p_{i'}^{(1)}$ is the probability that $\mathscr{F}_2$ gets the current relay and the reward offered by the next relay to $\mathscr{F}_1$ is $r_{i'}$. Similarly, $\nu_1 p_{j'}^{(2)}$ is the probability (again when the joint-action is $(\mathtt{s},\mathtt{s})$) that $\mathscr{F}_1$ gets the relay and the reward value of the next relay to $\mathscr{F}_2$ is $r_{j'}$.

Next, when the state is of the form $x = (r_i, \boldsymbol{t})$ (i.e., $\mathscr{F}_2$ has already terminated) the transition probabilities depend only on the action a_1 of $\mathscr{F}_1$ and is given by,

$$T(x'|x,a) = \begin{cases} p_{i'}^{(1)} & \text{if } a_1 = \mathtt{c}, x' = (r_{i'}, \boldsymbol{t}) \\ 1 & \text{if } a_1 = \mathtt{s}, x' = (\boldsymbol{t}, \boldsymbol{t}) \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Similarly one can write the expression for $T(x'|x,a)$ when the state is $x = (\boldsymbol{t}, r_j)$. Finally, the state $(\boldsymbol{t},\boldsymbol{t})$ is absorbing so that $T((\boldsymbol{t},\boldsymbol{t})|(\boldsymbol{t},\boldsymbol{t}), a) = 1$.

One-Step Costs: The one-step costs should be such that, for any policy pair (π_1, π_2), the sum of all one-step costs incurred by $\mathscr{F}_\rho$ ($\rho = 1,2$) should equal the total cost in (2). With this in mind, in Table 1 we write the pair of one-step costs, $(g_1(x,a), g_2(x,a))$, incurred by $\mathscr{F}_1$ and $\mathscr{F}_2$ for different joint-actions, $a = (a_1, a_2)$, when the current state is $x = (r_i, r_j)$.

$a = (a_1, a_2)$	$(g_1(x,a), g_2(x,a))$
$(\mathtt{c},\mathtt{c})$	(τ, τ)
$(\mathtt{c},\mathtt{s})$	$(\tau, -\eta_2 r_j)$
$(\mathtt{s},\mathtt{c})$	$(-\eta_1 r_i, \tau)$
$(\mathtt{s},\mathtt{s})$	$(-\eta_1 r_i, \tau)$ w.p. ν_1 $(\tau, -\eta_2 r_j)$ w.p. ν_2

TABLE 1
One-step costs when $x = (r_i, r_j)$.

From Table 1 we see that if the joint action is $(\mathtt{c},\mathtt{c})$ then both forwarders continue incurring a cost of τ which is the average time until the next relay arrives. When one of the forwarder, say $\mathscr{F}_2$, chooses to stop (i.e., the joint action is $(\mathtt{c},\mathtt{s})$) then $\mathscr{F}_2$, forwarding its packet to the chosen relay,

incurs a terminating cost of $-\eta_2 r_j$, while $\mathscr{F}_1$ simply continues incurring an average waiting delay of τ. Analogous is the case whenever the joint action is (s,c). Finally, if both forwarders compete (i.e., the case (s,s)), then with probability ν_ρ, $\mathscr{F}_\rho$ gets the relay incurring the terminating cost while the other forwarder has to continue.

a_1	$(g_1(x,a), g_2(x,a))$
c	$(\tau, 0)$
s	$(-\eta_1 r_i, 0)$

TABLE 2
$x = (r_i, \boldsymbol{t})$

a_2	$(g_1(x,a), g_2(x,a))$
c	$(0, \tau)$
s	$(0, -\eta_2 r_j)$

TABLE 3
$x = (\boldsymbol{t}, r_j)$

When the state is of the form $(r_i, \boldsymbol{t})$ the cost incurred by $\mathscr{F}_2$ is 0 for any joint-action a, and further the one-step cost incurred by $\mathscr{F}_1$ depends only on the action a_1 of $\mathscr{F}_1$. Analogous situation holds for $\mathscr{F}_2$ when the state is $(\boldsymbol{t}, r_j)$. These costs are given in Table 2 and 3, respectively. Finally, the cost incurred by both the forwarders once the termination state $(\boldsymbol{t},\boldsymbol{t})$ is reached is 0.

Now, given a policy pair (π_1, π_2) (recall Definition 1) and an initial state $x \in \mathcal{X}$, let $\{X_k : k \geq 1\}$ denote the sequence of (random) states traversed by the system, and let $\{(A_{1,k}, A_{2,k}) : k \geq 1\}$ denote the sequence of joint-actions. The total cost in (2) can now be expressed as the sum of all the one-step costs as follows:

$$J^{(\rho)}_{\pi_1,\pi_2}(x) \;=\; \sum_{k=1}^{\infty} \mathbb{E}^{x}_{\pi_1,\pi_2}\Big[g_\rho(X_k, (A_{1,k}, A_{2,k}))\Big]. \quad (9)$$

B. Characterization of NEPPs

States of the form $(r_i, \boldsymbol{t})$, $(\boldsymbol{t}, r_j)$

Once the system enters a state of the form $(r_i, \boldsymbol{t})$, since only $\mathscr{F}_1$ is present in the system, we essentially have an MDP problem where $\mathscr{F}_1$ is attempting to optimize its cost. Formally, if (π_1^*, π_2^*) is an NEPP then it can be argued[2] that $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, \boldsymbol{t})$ is the optimal cost to $\mathscr{F}_1$ with π_1^* being an optimal policy; the cost incurred by $\mathscr{F}_2$ is 0 and π_2^* can be arbitrary, but for simplicity we fix $\pi_2^*(r_i, \boldsymbol{t}) = \mathsf{s}$ for all $i \in [n]$. Hence $J^{(1)}_{\pi_1^*,\pi_2^*}(\cdot, \boldsymbol{t})$ satisfies the following Bellman optimality equation:

$$J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, \boldsymbol{t}) \;=\; \min\Big\{-\eta_1 r_i, D^{(1)}\Big\}, \quad (10)$$

where

$$D^{(1)} \;=\; \tau + \sum_{i'} p^{(1)}_{i'} J^{(1)}_{\pi_1^*,\pi_2^*}(r_{i'}, \boldsymbol{t}) \quad (11)$$

is the expected *cost of continuing alone in the system* (τ is the one-step cost and the remaining term is the future cost-to-go). $-\eta_1 r_i$ in the min-expression above is the cost of stopping. Thus, denoting $\frac{D^{(1)}}{-\eta_1}$ by $\alpha^{(1)}$, whenever the state is of the form $(r_i, \boldsymbol{t})$ an optimal policy is as follows:

$$\pi_1^*(r_i, \boldsymbol{t}) \;=\; \begin{cases} \mathsf{s} & \text{if } r_i \geq \alpha^{(1)} \\ \mathsf{c} & \text{otherwise.} \end{cases} \quad (12)$$

Remark: As mentioned earlier (recall the discussion in related work), the solution to the basic relay selection problem, comprising a single forwarder (say only $\mathscr{F}_1$) and a finite number of relays N, is characterized in terms of a single threshold α. Furthermore, from our earlier work [4, Section 6] we know that α is the unique fixed point of

$$\beta^{(1)}(x) = \mathbb{E}\Big[\max\{x, R_1\}\Big] - \frac{\tau}{\eta_1}, \quad (13)$$

where the expectation in the above expression is w.r.t. the p.m.f. $p^{(1)}$ of R_1. Here we will show that $\alpha^{(1)}$ is the fixed point of $\beta^{(1)}$, formalizing our earlier claim that the competitive model with only one forwarder and the infinite horizon basic model are equivalent. Although this result can be deduced by showing the equivalence between our competitive model with a single forwarder and the infinite horizon version of the asset selling problem, we prove it here for completeness.

Lemma 1: $\alpha^{(1)}$ is the unique fixed point of $\beta^{(1)}(x)$ ($x \in (-\infty, r_n]$) in (13).

Proof: We will first show that $\beta^{(1)}$ is a contraction mapping. Then, from the Banach fixed point theorem [38] it follows that there exists a unique fixed point α^* of $\beta^{(1)}$. Next, through an induction argument we will prove that $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, \boldsymbol{t}) = \min\Big\{-\eta_1 r_i, -\eta_1 \alpha^*\Big\}$. Finally, substituting for $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, \boldsymbol{t})$ in $\alpha^{(1)} = \frac{D^{(1)}}{-\eta_1}$ (recall $D^{(1)}$ from (11)) and simplifying, we obtain the desired result. Details of the proof are available in Appendix A. ■

Similarly, when the state is of the form $(\boldsymbol{t}, r_j)$ (i.e., $\mathscr{F}_1$ has already terminated), if (π_1^*, π_2^*) is an NEPP then, $J^{(1)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t}, r_j) = 0$ and $\pi_1^*(\boldsymbol{t}, r_j) = \mathsf{s}$, while $J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t}, r_j)$ satisfies

$$J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t}, r_j) \;=\; \min\Big\{-\eta_2 r_j, D^{(2)}\Big\}, \quad (14)$$

where $D^{(2)} = \tau + \sum_{j'} p^{(2)}_{j'} J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t}, r_{j'})$. Further, $\alpha^{(2)} = \frac{D^{(2)}}{-\eta_2}$, is the unique fixed point of $\beta^{(2)}(x) = \mathbb{E}\Big[\max\{x, R_2\}\Big] - \frac{\tau}{\eta_2}$, where now the expectation is w.r.t. the p.m.f. $p^{(2)}$ of R_2. Finally, an optimal policy π_2^* is such that

$$\pi_2^*(\boldsymbol{t}, r_j) \;=\; \begin{cases} \mathsf{s} & \text{if } r_j \geq \alpha^{(2)} \\ \mathsf{c} & \text{otherwise.} \end{cases} \quad (15)$$

States of the form (r_i, r_j)

This is the more interesting case where both forwarders are present in the system and are competing to choose a relay. When the state is of the form (r_i, r_j), if $\mathscr{F}_1$ decides to continue while $\mathscr{F}_2$ chooses to stop (i.e., the joint-action is (c,s)), then $\mathscr{F}_2$ terminates by incurring a cost of $-\eta_2 r_j$ so that the next state is of the form $(r_{i'}, \boldsymbol{t})$. Hence the expected total cost incurred by $\mathscr{F}_1$, if it uses the policy in (12) from the next stage onwards, is $D^{(1)}$ (recall (11)). Similarly, if the joint-action is (s,c) then $\mathscr{F}_1$ terminates incurring a cost of $-\eta_1 r_i$, and $\mathscr{F}_2$ incurs a cost of $D^{(2)}$ if it uses the policy in (15) from the next stage onwards.

If both forwarders decide to stop (joint-action is (s,s)) then with probability ν_1, $\mathscr{F}_1$ gets the relay in which case $\mathscr{F}_2$ continues alone, and with probability ν_2 it is vice versa. Thus, the expected cost incurred by $\mathscr{F}_1$ is,

$$E^{(1)}(r_i) \;=\; \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}, \quad (16)$$

[2]Using the definition of an NEPP and the fact that the costs and the state transitions do not depend on the policy of the other forwarder anymore.

and that by $\mathscr{F}_2$ is,

$$E^{(2)}(r_j) \;=\; \nu_1 D^{(2)} + \nu_2(-\eta_2 r_j). \qquad (17)$$

Finally, if both forwarders choose to continue (i.e., if the joint-action is (c,c)) then the next state is again of the form $(r_{i'}, r_{j'})$. Thus if (π_1, π_2) is the policy pair used from the next stage onwards then the expected costs incurred by $\mathscr{F}_1$ and $\mathscr{F}_2$ are, respectively,

$$C^{(1)}_{\pi_1,\pi_2} \;=\; \tau + \sum_{i',j'} p_{i',j'} J^{(1)}_{\pi_1,\pi_2}(r_{i'}, r_{j'}) \qquad (18)$$

$$C^{(2)}_{\pi_1,\pi_2} \;=\; \tau + \sum_{i',j'} p_{i',j'} J^{(2)}_{\pi_1,\pi_2}(r_{i'}, r_{j'}). \qquad (19)$$

We are now ready to state the following main theorem (adapted from [29]), which relates the "NEPPs of the stochastic game" with the "Nash equilibrium strategies of a certain static game" played at a stage. The various cost terms described above are used to construct this static game. We state the theorem below with the understanding that for states of the form $x = (r_i, \boldsymbol{t})$ and $x = (\boldsymbol{t}, r_j)$, $\pi_1^*(x)$ and $\pi_2^*(x)$ are as in (12) and (15), respectively.

Theorem 1: Given a policy pair, (π_1^*, π_2^*), for each state $x = (r_i, r_j)$ construct the static game given in Table 4.

	c	s
c	$C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*}$	$D^{(1)}, -\eta_2 r_j$
s	$-\eta_1 r_i, D^{(2)}$	$E^{(1)}(r_i), E^{(2)}(r_j)$

TABLE 4
STATIC STAGE GAME.

Then the following statements are equivalent:

a) (π_1^*, π_2^*) is an NEPP.
b) For each $x = (r_i, r_j)$, $(\pi_1^*(x), \pi_2^*(x))$ is a *Nash equilibrium (NE) strategy* for the game in Table 4. Further, the expected payoff pair at this NE strategy is, $(J^{(1)}_{\pi_1^*,\pi_2^*}(x), J^{(2)}_{\pi_1^*,\pi_2^*}(x))$.

Proof: Although the proof of this theorem is along the lines of the proof of Theorem 4.6.5 in [29], however some additional efforts are required since the proof in [29] is for the case where the costs are discounted, while ours is a total cost undiscounted stochastic game. Further, the presence of a cost-free absorption state for each player renders our problem *transient* by which we mean, when the policy of one player is fixed the problem of obtaining the optimal policy for the other player is a stopping problem [39]. Using this property we have modified the proof of [29, Theorem 4.6.5] appropriately so that the result holds for our case. For details, see Appendix B. ■

Discussion: In this discussion for simplicity we will omit (π_1^*, π_2^*) from all the associated notations. Now, Theorem 1 can be seen as an extension of the Bellman optimality equation in (10), where to obtain $J^{(1)}(r_i, \boldsymbol{t})$ we require the cost term $D^{(1)}$ in (11), which in turn depends on the function $J^{(1)}(\cdot, \boldsymbol{t})$. This essentially suggests that $J^{(1)}(\cdot, \boldsymbol{t})$ is the fixed point of the Bellman equation in (10). Similarly, here we see that, given the cost pair $(C^{(1)}, C^{(2)})$, one can obtain $(J^{(1)}(x), J^{(2)}(x))$ by solving the game in Table 4. However, computing $(C^{(1)}, C^{(2)})$ itself will require the function pair $(J^{(1)}(\cdot), J^{(2)}(\cdot))$, thus suggesting that $(J^{(1)}(\cdot), J^{(2)}(\cdot))$ has to be fixed point of a

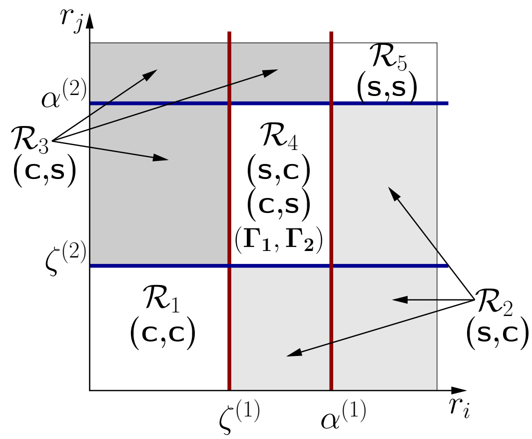


Fig. 2. Illustration of the various regions along with the NE strategies corresponding to these regions.

mapping which involves computing the payoff pair of the static game in Table 4. Furthermore, analogous to computing the minimum in (10) to obtain the optimal action, here, by computing the NE strategies of the game in Table 4 we obtain the solution to our stochastic game.

Assuming that the cost pair $(C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*})$ is given to us, we now proceed to obtain all the NE strategies of the game in Table 4. We will first require the following key lemma.

Lemma 2: For an NEPP, (π_1^*, π_2^*), the various costs are ordered as follows:

$$D^{(1)} \leq C^{(1)}_{\pi_1^*,\pi_2^*} \text{ and } D^{(2)} \leq C^{(2)}_{\pi_1^*,\pi_2^*}. \qquad (20)$$

Proof: See Appendix C. ■

Discussion: The above lemma becomes intuitive once we recall that $D^{(1)}$ is the optimal cost incurred by $\mathscr{F}_1$ if it is alone in the system, while $C^{(1)}_{\pi_1^*,\pi_2^*}$ is the cost incurred if $\mathscr{F}_2$ is also present, and competing with $\mathscr{F}_1$ in choosing a relay. One would expect $\mathscr{F}_1$ to incur a lower cost without the competing forwarder.

For notational simplicity, from here on, we will denote the costs $C^{(1)}_{\pi_1^*,\pi_2^*}$ and $C^{(2)}_{\pi_1^*,\pi_2^*}$ as simply $C^{(1)}$ and $C^{(2)}$. We will write $\mathbf{C}$ for the pair $(C^{(1)}, C^{(2)})$. An important consequence of Lemma 2 is that, while solving the game in Table 4, it is sufficient to only consider cost pairs, $(C^{(1)}, C^{(2)})$, which are ordered as in the lemma; the other cases (e.g., $D^{(1)} > C^{(1)}$ or $D^{(2)} > C^{(2)}$) cannot occur, and hence need not be considered. Further, for convenience let us denote the thresholds $\frac{C^{(1)}}{-\eta_1}$ and $\frac{C^{(2)}}{-\eta_2}$ by $\zeta^{(1)}$ and $\zeta^{(2)}$, respectively (recall that we already have, $\alpha^{(1)} = \frac{D^{(1)}}{-\eta_1}$ and $\alpha^{(2)} = \frac{D^{(2)}}{-\eta_2}$). Then, the solution (i.e., the NE strategies) to the game in Table 4, for each (r_i, r_j) pair, is as depicted in Fig. 2.

We see that the thresholds $(\alpha^{(1)}, \zeta^{(1)})$ and $(\alpha^{(2)}, \zeta^{(2)})$ partition the reward pair set, $\{(r_i, r_j) : i, j \in [n]\}$, into 5 regions $(\mathcal{R}_1, \cdots, \mathcal{R}_5)$[3] such that the NE strategy (strategies) corresponding to each region are different. For instance, for any $(r_i, r_j) \in \mathcal{R}_1$, (c,c) (i.e., both forwarders continue) is the only NE strategy, while within $\mathcal{R}_2$, (s,c) is the NE strategy, and so on. All regions contain a unique pure NE strategy

[3]These regions depend on the cost pair $\mathbf{C}$; for simplicity we neglect $\mathbf{C}$ in their notation. However, we will invoke this dependency when required.

except for $\mathcal{R}_4$ where $(\mathtt{s},\mathtt{c})$, $(\mathtt{c},\mathtt{s})$, and the mixed strategy (Γ_1,Γ_2) (Γ_ρ is the probability with which $\mathscr{F}_\rho$ chooses $\mathtt{s}$) are all NE strategies. The expression for Γ_1 is

$$\Gamma_1 = \frac{-\eta_2 r_j - C^{(2)}}{\Big(-\eta_2 r_j - C^{(2)}\Big) - \Big(E^{(2)}(r_j) - D^{(2)}\Big)}. \tag{23}$$

Analogously one can write the expression for Γ_2. For details on how to solve the game in Table 4 to obtain the various regions, see Appendix D. Finally, we summarize the observations made thus far in the form of the following theorem.

Theorem 2: The NE strategies of the game in Table 4 are completely characterized by the threshold pairs $(\alpha^{(\rho)},\zeta^{(\rho)})$, $\rho = 1,2$ as follows (recall Fig. 2 for illustration):

- If r_i is less than $\zeta^{(1)}$, then the NE strategy recommends $\mathtt{c}$ for $\mathscr{F}_1$ irrespective of the reward value r_j of $\mathscr{F}_2$.
- On the other hand, if r_i is more than $\alpha^{(1)}$, then the NE strategy recommends action $\mathtt{s}$ for $\mathscr{F}_1$ irrespective of the value of r_j (note that this is exactly the action $\mathscr{F}_1$ would choose if it was alone in the system; see the discussion following (12)).
- Finally, the presence of the competing forwarder $\mathscr{F}_2$ is felt by $\mathscr{F}_1$ only when its reward value r_i is between $\zeta^{(1)}$ and $\alpha^{(1)}$, in which case the NE strategies are: $(\mathtt{s},\mathtt{c})$ if $r_j < \zeta^{(2)}$; $(\mathtt{s},\mathtt{c})$, $(\mathtt{c},\mathtt{s})$ and (Γ_1,Γ_2) if $\zeta^{(2)} \le r_j \le \alpha^{(2)}$; and $(\mathtt{c},\mathtt{s})$ if $r_j > \alpha^{(2)}$.

Analogous results hold for $\mathscr{F}_2$.

C. Constructing NEPPs from NE strategies

The cost terms $D^{(1)}$ and $D^{(2)}$ can be easily computed by solving the optimality equations (10) and (14), respectively. Alternatively, we can first compute the fixed points of $\beta^{(1)}(\cdot)$ and $\beta^{(2)}(\cdot)$ to obtain $\alpha^{(1)}$ and $\alpha^{(2)}$, respectively (recall Lemma 1). Then, $D^{(1)} = -\eta_1\alpha^{(1)}$ and $D^{(2)} = -\eta_2\alpha^{(2)}$.

The costs $C^{(1)}$ and $C^{(2)}$ (in (18) and (19)) depend on the particular NEPP used, i.e., require the cost terms $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,r_j)$ and $J^{(2)}_{\pi_1^*,\pi_2^*}(r_i,r_j)$ for all (r_i,r_j) to compute them. Conversely, Part-(b) of Theorem 1 suggests that $J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,r_j)$ (respectively, $J^{(2)}_{\pi_1^*,\pi_2^*}(r_i,r_j)$) can be obtained by computing the expected cost incurred by $\mathscr{F}_1$ (respectively, $\mathscr{F}_2$) at a NE strategy of the game in Table 4, which in turn requires the terms $C^{(1)}$ and $C^{(2)}$. Hence, to obtain $(C^{(1)},C^{(2)})$ we proceed by expressing $(C^{(1)},C^{(2)})$ as the fixed point of a mapping $\mathcal{T}$ which can then be used to compute these costs.

Suppose (π_1^*,π_2^*) is a NEPP such that for all $x = (r_i,r_j) \in \mathcal{R}_4(\mathbf{C})$ the NE strategy $(\pi_1^*(x),\pi_2^*(x))$ is $(\mathtt{s},\mathtt{c})$. Then using part 2(b) of Theorem 1 we can write,

$$J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,r_j) = \begin{cases} C^{(1)} & \text{if } (r_i,r_j)\in\mathcal{R}_1(\mathbf{C}) \\ -\eta_1 r_i & \text{if } (r_i,r_j)\in\mathcal{R}_2(\mathbf{C}) \\ D^{(1)} & \text{if } (r_i,r_j)\in\mathcal{R}_3(\mathbf{C}) \\ -\eta_1 r_i & \text{if } (r_i,r_j)\in\mathcal{R}_5(\mathbf{C}) \\ E^{(1)}(r_i) & \text{if } (r_i,r_j)\in\mathcal{R}_4(\mathbf{C}). \end{cases} \tag{24}$$

Using the above in (18), $C^{(1)}$ can be written as $C^{(1)} = \mathcal{T}_1(\mathbf{C})$ where the function $\mathcal{T}_1(\mathbf{C})$ is as in (21) (where for simplicity, we have used (i,j) instead of (r_i,r_j)). Similarly, $C^{(2)}$ can be expressed as $C^{(2)} = \mathcal{T}_2(\mathbf{C})$; see (22). Thus, $\mathbf{C}$ is a fixed point of the mapping $\mathcal{T}(\mathbf{C}) := (\mathcal{T}_1(\mathbf{C}),\mathcal{T}_2(\mathbf{C}))$.

We do not have results showing that $\mathcal{T}$ indeed has a fixed point or equivalently that an NEPP (π_1^*,π_2^*) always exists,[4] although such a result holds for the discounted stochastic game [29, Theorem 4.6.4] (recall that ours is a transient stochastic game). However, in our numerical results section (Section VIII) we were able to numerically obtain $\mathbf{C}$ by iteration. Thus, we begin with an initial $\mathbf{C}(0)$ such that $C^{(1)}(0) < D^{(1)}$ and $C^{(2)}(0) < D^{(2)}$, and inductively iterate to obtain $\mathbf{C}(k) = \mathcal{T}(\mathbf{C}(k-1))$ until convergence is achieved. Finally, given a fixed point $\mathbf{C}$, we obtain the corresponding NEPP (π_1^*,π_2^*) by constructing the various regions as in Fig. 2.

Other NEPPs: Recall that to obtain $(C^{(1)},C^{(2)})$ we had restricted (π_1^*,π_2^*) to use NE strategy $(\mathtt{s},\mathtt{c})$ whenever $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C})$. We can similarly obtain NEPPs $(\pi_1^\circ,\pi_2^\circ)$ and $(\pi_1^\square,\pi_2^\square)$ (whose corresponding cost pairs are $\mathbf{C}_\circ$ and $\mathbf{C}_\square$) by restricting to the NE strategies $(\mathtt{c},\mathtt{s})$ and (Γ_1,Γ_2) whenever $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C}_\circ)$ and $(r_i,r_j)\in\mathcal{R}_4(\mathbf{C}_\square)$, respectively. In Section VIII we will numerically compare the performances of all these various NEPPs.

VI. Partially Observable Case

Let us first formally introduce a finite *location set* $\mathcal{L}$. Let L_k denote the *location* of the k-th relay. The locations $\{L_k : k \ge 1\}$ are i.i.d with their common p.m.f. being $(q_\ell : \ell\in\mathcal{L})$. Recall that for the PO case we assume that only $R_{\rho,k}$ is revealed to $\mathscr{F}_\rho$ $(\rho = 1,2)$. In addition, we will assume that L_k is revealed to both the forwarders.

Recalling the geographical forwarding example from Section III, the PO case corresponds to the scenario where, in addition to L_k, the gains $G_{\rho,k}$ are required to compute $R_{\rho,k}$, i.e., if $a < 1$ in (6). Hence, $\mathscr{F}_1$ not knowing $G_{2,k}$ cannot compute $R_{2,k}$. However, knowing the channel gain distribution (recall that the gains are identically distributed) it is possible

[4] This equivalence can be easily shown by first using (π_1^*,π_2^*) in part-(a) of Theorem 1 to conclude that part-(b) holds, and then simply from the definition of $\mathcal{T}$ it will follow that it has a fixed point. For the other direction, given a fixed point $\mathbf{C}$ of $\mathcal{T}$, one can easily obtain the corresponding NEPP (π_1^*,π_2^*) by constructing the various regions as shown in Fig. 2.

$$\mathcal{T}_1(\mathbf{C}) = \tau + \sum_{(i,j)\in\mathcal{R}_1(\mathbf{C})} p_{i,j}C^{(1)} + \sum_{(i,j)\in\mathcal{R}_2(\mathbf{C})\cup\mathcal{R}_4(\mathbf{C})} p_{i,j}(-\eta_1 r_i) + \sum_{(i,j)\in\mathcal{R}_3(\mathbf{C})} p_{i,j}D^{(1)} + \sum_{(i,j)\in\mathcal{R}_5(\mathbf{C})} p_{i,j}E^{(1)}(r_i) \tag{21}$$

$$\mathcal{T}_2(\mathbf{C}) = \tau + \sum_{(i,j)\in\mathcal{R}_1(\mathbf{C})} p_{i,j}C^{(2)} + \sum_{(i,j)\in\mathcal{R}_2(\mathbf{C})\cup\mathcal{R}_4(\mathbf{C})} p_{i,j}D^{(2)} + \sum_{(i,j)\in\mathcal{R}_3(\mathbf{C})} p_{i,j}(-\eta_2 r_j) + \sum_{(i,j)\in\mathcal{R}_5(\mathbf{C})} p_{i,j}E^{(2)}(r_j) \tag{22}$$

for $\mathscr{F}_1$ to compute the probability distribution of $R_{2,k}$ given L_k. Similarly, $\mathscr{F}_2$ can compute the distribution of $R_{1,k}$ given L_k. Further, since the gains, $(G_{1,k}, G_{2,k})$, are independent, it follows that $R_{1,k}$ and $R_{2,k}$ are independent given L_k (but unconditionally they may be dependent).

Formally, given that $L_k = \ell$, we will assume the following *independence condition*:

$$p_{R_1,R_2|L_k}(r_i, r_j|\ell) = p_{R_1|L_k}(r_i|\ell)p_{R_2|L_k}(r_j|\ell). \tag{25}$$

For simplicity, we will denote the conditional p.m.f.s $p_{R_1|L_k}(r_i|\ell)$ and $p_{R_2|L_k}(r_j|\ell)$, $i,j \in [n]$, by $p^{(1)}_{i|\ell}$ and $p^{(2)}_{j|\ell}$, respectively.

Remark: Usually for a model with partial observations the belief that $\mathscr{F}_1$ will maintain about $R_{2,k}$ will simply be the conditional distribution $p_{R_2|R_1}(r_j|r_i) = \frac{p_{i,j}}{p^{(1)}_i}$. However, we have exploited the particular structure in our reward expression to come up with the independence condition in (25). This condition will enable us to prove a key result later which is otherwise not possible (see the remark following Lemma 4). Finally, all our subsequent results will hold for a more general model wherever the independence condition in (25) will hold.

We will now proceed to formulate our partially observable model as a partially observable stochastic game (POSG). We will first formally describe the problem setting and then briefly discuss POSGs, before proceeding to our main results.

A. Problem Formulation

The actual state space of the system continues to be $\mathcal{X}$ (see (1)). However, each forwarder now gets to observe only its part of the actual state (i.e., only its reward value) along with the relay's location. Thus, when the k-th relay arrives, and if both forwarders are still competing then the observations of $\mathscr{F}_1$ and $\mathscr{F}_2$ are of the form (r_i, ℓ) and (ℓ, r_j), respectively, where (r_i, r_j) is the actual state, $L_k = \ell$ is the location of the k-th relay. Suppose $\mathscr{F}_2$ has already terminated before stage k then[5] the location information is no more required by $\mathscr{F}_1$, and hence we will denote its observation as $(r_i, \boldsymbol{t})$ which is simply the system state. Finally, when $\mathscr{F}_1$ terminates we use $\boldsymbol{t}$ to denote its subsequent observations. Thus, we can write the *observation space* of $\mathscr{F}_1$ as,

$$\mathcal{O}_1 \quad = \quad \Big\{(r_i, \ell), (r_i, \boldsymbol{t}), \boldsymbol{t} : i \in [n], \ell \in [m]\Big\}. \tag{26}$$

Similarly, the observation space of $\mathscr{F}_2$ is given by

$$\mathcal{O}_2 \quad = \quad \Big\{(\ell, r_j), (\boldsymbol{t}, r_j), \boldsymbol{t} : j \in [n], \ell \in [m]\Big\}. \tag{27}$$

Definition 3: We will modify[6] the definition of a policy pair, $(\overline{\pi}_1, \overline{\pi}_2)$ (see Definition 1), such that $\overline{\pi}_1 : \mathcal{O}_1 \rightarrow \{\mathtt{s}, \mathtt{c}\}$ and $\overline{\pi}_2 : \mathcal{O}_2 \rightarrow \{\mathtt{s}, \mathtt{c}\}$. Thus, the decision to stop or continue by $\mathscr{F}_1$ and $\mathscr{F}_2$, when the k-th relay arrives is based on their respective observations $o_{1,k} \in \mathcal{O}_1$ and $o_{2,k} \in \mathcal{O}_2$.

Remark: Note that we have restricted the PO policies to be deterministic (and as before stationary), i.e., $\overline{\pi}_1(o_1)$ is either $\mathtt{s}$ or $\mathtt{c}$ without mixing between the two. Let Π_D denote the set of all such deterministic policies. Restricting to Π_D is primarily to simplify the analysis. However, it is not immediately clear if a partially observable NEPP (to be formally defined very soon) should even exist within the class Π_D. Our main result is to construct a Bayesian stage game and prove that this game contains pure strategy (or deterministic) NE vectors using which PO-NEPPs in Π_D can be constructed.

Let $\{(O_{1,k}, O_{2,k})\colon k \geq 1\}$, denote the sequence of joint-observation at stage k, and let $\{X_k : k \geq 1\}$ as before denote the sequence of states. Then the expected cost incurred by $\mathscr{F}_\rho$, $\rho = 1,2$, when the PO policy pair used is $(\overline{\pi}_1, \overline{\pi}_2)$, and when its initial observation is o_ρ, can be written as

$$G^{(\rho)}_{\overline{\pi}_1,\overline{\pi}_2}(o_\rho) = \sum_{k=1}^{\infty} \mathbb{E}^{o_\rho}_{\overline{\pi}_1,\overline{\pi}_2}\Big[g_\rho(X_k, (A_{1,k}, A_{2,k}))\Big], \tag{28}$$

where $A_{1,k} = \overline{\pi}_1(O_{1,k})$ and $A_{2,k} = \overline{\pi}_2(O_{2,k})$.

Similar to the completely observable case, the objective for the partially observable (PO) case is to characterize PO-NEPPs which are defined as follows:

Definition 4: We say that a PO policy pair $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ is a PO-NEPP if $G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(o_1) \leq G^{(1)}_{\overline{\pi}_1,\overline{\pi}^*_2}(o_1)$ for all $o_1 \in \mathcal{O}_1$ and PO policy $\overline{\pi}_1 \in \Pi_D$, and $G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(o_2) \leq G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}_2}(o_2)$ where $o_2 \in \mathcal{O}_2$ and $\overline{\pi}_2 \in \Pi_D$.

We will end this section with the expressions for the various cost terms corresponding to a PO-NEPP, which are analogues of the cost terms in Section V.

Various Cost Terms: Recall the expression for $D^{(1)}$ from (11). Given a NEPP (π^*_1, π^*_2), $D^{(1)}$ is the cost incurred by $\mathscr{F}_1$ if it continues alone. Similar expressions can be written for a PO-NEPP $(\overline{\pi}^*_1, \overline{\pi}^*_2)$:

$$\overline{D}^{(1)} \quad = \quad \tau + \sum_{i'} p^{(1)}_{i'} G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_{i'}, \boldsymbol{t}). \tag{29}$$

Similarly, for $\mathscr{F}_2$, the cost of continuing alone is

$$\overline{D}^{(2)} \quad = \quad \tau + \sum_{j'} p^{(2)}_{j'} G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(\boldsymbol{t}, r_{j'}). \tag{30}$$

The following lemma will be useful.

Lemma 3: Let (π^*_1, π^*_2) be an NEPP and $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ be a PO-NEPP then $J^{(1)}_{\pi^*_1,\pi^*_2}(r_i, \boldsymbol{t}) = G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i, \boldsymbol{t})$ and $J^{(2)}_{\pi^*_1,\pi^*_2}(\boldsymbol{t}, r_j) = G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(\boldsymbol{t}, r_j)$.

Proof: Whenever $\mathscr{F}_1$ is alone in the system, all its observations (which are of the form $(r_i, \boldsymbol{t})$ until $\mathscr{F}_1$ terminates) are exactly the actual states traversed by the system. Hence the problem of obtaining $G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i, \boldsymbol{t})$ is identical to the MDP problem of obtaining $J^{(1)}_{\pi^*_1,\pi^*_2}(r_i, \boldsymbol{t})$ in Section V-B, so that $G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i, \boldsymbol{t})$ satisfies the Bellman equation in (10). Since the solution to (10) is unique [39] we obtain $J^{(1)}_{\pi^*_1,\pi^*_2}(r_i, \boldsymbol{t}) = G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i, \boldsymbol{t})$. Similarly it follows that $J^{(2)}_{\pi^*_1,\pi^*_2}(\boldsymbol{t}, r_j) = G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(\boldsymbol{t}, r_j)$. ■

Discussion: An immediate consequence of the above lemma is that $\overline{D}^{(1)} = D^{(1)}$ and $\overline{D}^{(2)} = D^{(2)}$. Further, if $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ is a PO-NEPP then for states of the form $(r_i, \boldsymbol{t})$, $\overline{\pi}^*_1(r_i, \boldsymbol{t})$ is same as $\pi^*_1(r_i, \boldsymbol{t})$ in (12). Similarly, for states of the form $(\boldsymbol{t}, r_j)$, $\overline{\pi}^*_2(\boldsymbol{t}, r_j)$ is same as that in (15).

[5] As mentioned earlier, $\mathscr{F}_1$ will come to know about $\mathscr{F}_2$'s termination by listening to the exchange of control packets between $\mathscr{F}_2$ and the chosen relay just before termination.

[6] In this section we will apply overline to most of the symbols in order to distinguish them from the corresponding symbols that have already appeared in Section V.

However, the analogues of the cost terms $C^{(1)}_{\pi_1,\pi_2}$ and $C^{(2)}_{\pi_1,\pi_2}$ (recall (18) and (19)) are different for the partially observable case. The expressions for these are,

$$\overline{C}^{(1)}_{\overline{\pi}_1,\overline{\pi}_2} = \tau + \sum_{\ell',i'} q_{\ell'} \cdot p^{(1)}_{i'|\ell'} \cdot G^{(1)}_{\overline{\pi}_1,\overline{\pi}_2}(r_{i'},\ell'), \quad (31)$$

$$\overline{C}^{(2)}_{\overline{\pi}_1,\overline{\pi}_2} = \tau + \sum_{\ell',j'} q_{\ell'} \cdot p^{(2)}_{j'|\ell'} \cdot G^{(2)}_{\overline{\pi}_1,\overline{\pi}_2}(\ell',r_{j'}). \quad (32)$$

Finally, similar to the result in Lemma 2, we can show that for a PO-NEPP $(\overline{\pi}_1^*,\overline{\pi}_2^*)$,

$$\overline{D}^{(1)} \leq \overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*} \text{ and } \overline{D}^{(2)} \leq \overline{C}^{(2)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}. \quad (33)$$

The proof of these is along exactly the same lines as the proof of Lemma 2. We do not repeat it for brevity.

B. Partially Observable Stochastic Game (POSG)

A POSG is a tuple $(\mathcal{N},\mathcal{X},\mathcal{O},\{\mathcal{A}_\rho\},\hat{T},\{g_\rho\})$, where $\mathcal{N}$, $\mathcal{X}$, $\mathcal{A}_\rho$, and g_ρ are as before (see Section V-A), while

- $\mathcal{O} = \times_{\rho\in\mathcal{N}}\mathcal{O}_\rho$ is the joint-observation space, with $\mathcal{O}_\rho$ being the observation space of player ρ, and
- $\hat{T} : \mathcal{X}\times\mathcal{O}\times\mathcal{A} \to \Delta(\mathcal{X}\times\mathcal{O})$ is the transition function where $\hat{T}(x',o'|x,o,a)$ is the probability that the next state and the joint-observation is (x',o') conditioned on the event that the current state, joint-observation and joint-action is (x,o,a).

In the previous section we have seen that the NEPPs for a stochastic game can be obtained by constructing a normal-form static stage game. Similarly for POSGs, there is work (for instance see, [40]) that constructs a game which is effectively played at each stage, however, with the players not knowing the exact state of the system the stage game now happens to be a *Bayesian game* [41, Chapter 9]. Hence, the drawback with POSGs in general is that, at each stage k, each player needs to maintain a belief (distribution) about the entire history of joint-observations and joint-actions,
$((o_{1,1},o_{2,1}),(a_{1,1},a_{1,2}),\cdots,(a_{1,k-1},a_{2,k-1}),(o_{1,k},o_{2,k})$,
(referred to as the *joint-type of the Bayesian game*), obtaining which for a general POSG is computationally intensive.

For this reason the authors in [42] have studied a restriction of POSGs referred to as, Markov games of Incomplete information (MGII). In MGIIs the transition function $\hat{T}$ satisfies the following Markov property: player-1's belief about the player-2's current observation, o_2', is independent of player-2's previous observation, o_2, given the current state, x', previous state, x, and player-1's current and previous observations, o'_1 and o_1, respectively, i.e., for two different observations $u,v \in \mathcal{O}_2$ of player-2, $\hat{T}(o'_2|x',x,o'_2,o_1,o_2=u) = \hat{T}(o'_2|x',x,o'_1,o_1,o_2=v)$. Similar Markov structure should hold for other players also. For our case it is easy to check that the above condition is trivially satisfied, primarily because all the associated random variables, $\{L_k\}$ and $\{(R_{1,k},R_{2,k})\}$, are i.i.d. across the stage index k.

A major advantage with MGIIs is that the *current joint-observation constitutes the type of the Bayesian game to be played at that stage*. With this in mind, we will set up a Bayesian stage game in the next section, with (r_i,ℓ) and (ℓ,r_j) constituting the type of the game at stage k, provided both forwarders are still competing[7] at stage k.

C. Bayesian Stage Game

We are now ready to provide a solution to the partially observable case in terms of a certain Bayesian game [41, Chapter 9] which is effectively played at any stage whenever both forwarders are contending. For the completely observable case, given a policy pair (π_1,π_2), corresponding to each (r_i,r_j) pair we constructed the normal-form game in Table 4. However here, given a PO policy pair (π_1,π_2) and given the observation (r_i,ℓ), $\mathscr{F}_1$'s belief that the game in Table 4 (with $(C^{(1)}_{\pi_1,\pi_2},C^{(2)}_{\pi_1,\pi_2})$ replaced by $(\overline{C}^{(1)}_{\overline{\pi}_1,\overline{\pi}_2}\overline{C}^{(2)}_{\overline{\pi}_1,\overline{\pi}_2})$) will be played is $p^{(2)}_{j|\ell}$, $j\in[n]$. Hence, $\mathscr{F}_1$ needs to first compute the costs incurred for playing $\mathtt{s}$ and $\mathtt{c}$, averaged over all observations (ℓ,r_j), $j\in[n]$, of $\mathscr{F}_2$. We will formally develop these in the following.

Strategy vectors and corresponding costs: Fixing the PO-policy pair to be $(\overline{\pi}_1,\overline{\pi}_2)$ (unless otherwise stated), we will refer to the subsequent development (which includes, the strategy vectors, various costs, best responses and NE vectors, to be discussed next) as the *Bayesian game corresponding to* $(\overline{\pi}_1,\overline{\pi}_2)$, denoted $\mathcal{G}(\overline{\pi}_1,\overline{\pi}_2)$.

Definition 5: For $\ell\in\mathcal{L}$ (recall that $\mathcal{L}$ is the set of possible relay locations), we define a *strategy vector*, f_ℓ, of $\mathscr{F}_1$ as $f_\ell : \{r_i : i\in[n]\} \to \{\mathtt{s},\mathtt{c}\}$. Similarly, a strategy vector g_ℓ of $\mathscr{F}_2$ is $g_\ell : \{r_j : j\in[n]\} \to \{\mathtt{s},\mathtt{c}\}$. Thus, given the observation (r_i,ℓ) of $\mathscr{F}_1$, f_ℓ decides for $\mathscr{F}_1$ whether to stop or continue.

Now, given the strategy vector g_ℓ of $\mathscr{F}_2$, and the location information ℓ, $\mathscr{F}_1$'s belief that $\mathscr{F}_2$ will choose action $\mathtt{c}$ is

$$\widetilde{g}_\ell = \sum_{j:g_\ell(r_j)=\mathtt{c}} p^{(2)}_{j|\ell}; \quad (34)$$

$(1-\widetilde{g}_\ell)$ is the probability that $\mathscr{F}_2$ will stop. Thus, the expected cost incurred by $\mathscr{F}_1$ for playing $\mathtt{s}$ when its observation is (r_i,ℓ) and when $\mathscr{F}_2$ uses g_ℓ is

$$C^{(1)}_{\mathtt{s},g_\ell}(r_i) = \widetilde{g}_\ell(-\eta_1 r_i) + (1-\widetilde{g}_\ell)E^{(1)}(r_i), \quad (35)$$

where, recall from (16) that $E^{(1)}(r_i) = \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}$. The various terms in (35) can be understood as follows: $\widetilde{g}_\ell$ is the probability that $\mathscr{F}_2$ will continue in which case $\mathscr{F}_1$ (having chosen the action $\mathtt{s}$) stops, incurring a terminating cost of $-\eta_1 r_i$, while $(1-\widetilde{g}_\ell)$ is the probability that $\mathscr{F}_2$ will stop in which case the expected cost is, $\nu_1(-\eta_1 r_i)+\nu_2 D^{(1)}$; ν_1 is the probability that $\mathscr{F}_1$ gets the relay and terminates incurring a cost of $(-\eta_1 r_i)$, otherwise w.p. ν_2, $\mathscr{F}_2$ gets the relay in which case $\mathscr{F}_1$ continues alone, the expected cost of which is $\overline{D}^{(1)} = D^{(1)}$ (from Lemma 3).

The expected cost of continuing when $\mathscr{F}_1$'s observation is (r_i,ℓ) is

$$C^{(1)}_{\mathtt{c},g_\ell}(r_i) = \widetilde{g}_\ell \overline{C}^{(1)}_{\overline{\pi}_1,\overline{\pi}_2} + (1-\widetilde{g}_\ell)D^{(1)}. \quad (36)$$

[7] When only one forwarder is present we already know that the solution can be obtained by solving an MDP problem as in Section V-B (see Lemma 3).

From the above expression we see that the cost of continuing is a constant in the sense that it does not depend on the value of r_i. Hence we will denote it as simply $C^{(1)}_{\mathsf{c},g_\ell}$. Further, note that $C^{(1)}_{\mathsf{c},g_\ell}$ depends on the PO policy pair $(\overline{\pi}_1, \overline{\pi}_2)$, but for simplicity we have not shown this dependence in the notation for $C^{(1)}_{\mathsf{c},g_\ell}$.

Similarly for $\mathscr{F}_2$, when its observation is (ℓ, r_j) and when $\mathscr{F}_1$ uses f_ℓ, we have

$$\begin{aligned} C^{(2)}_{\mathsf{s},f_\ell}(r_j) &= \widetilde{f}_\ell(-\eta_2 r_j) + (1-\widetilde{f}_\ell)E^{(2)}(r_j) \\ C^{(2)}_{\mathsf{c},f_\ell} &= \widetilde{f}_\ell \overline{c}^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2} + (1-\widetilde{f}_\ell)D^{(2)}, \end{aligned}$$

where $\widetilde{f}_\ell = \sum_{i:f_\ell(r_i)=\mathsf{c}} p^{(1)}_{i|\ell}$.

Definition 6: We say that f_ℓ is the *best response vector of* $\mathscr{F}_1$ against the strategy vector g_ℓ played by $\mathscr{F}_2$, denoted $f_\ell = BR_1(g_\ell)$, if $f_\ell(r_i) = \mathsf{s}$ iff $C^{(1)}_{\mathsf{s},g_\ell}(r_i) \le C^{(1)}_{\mathsf{c},g_\ell}$. Note that such an f_ℓ is unique. Similarly, g_ℓ is the (unique) best response against f_ℓ if, $g_\ell(r_j) = \mathsf{s}$ iff $C^{(2)}_{\mathsf{s},f_\ell}(r_j) \le C^{(2)}_{\mathsf{c},f_\ell}$. We denote this as $g_\ell = BR_2(f_\ell)$.

Definition 7: For $\ell \in \mathcal{L}$, a pair of strategy vectors (f^*_ℓ, g^*_ℓ) is said to be a *Nash equilibrium (NE) vector* for the game $\mathcal{G}(\overline{\pi}_1, \overline{\pi}_2)$ iff $f^*_\ell = BR_1(g^*_\ell)$, and $g^*_\ell = BR_2(f^*_\ell)$.

As remarked earlier, it is not immediately clear whether a NE vector should even exist among the pure strategies for the game $\mathcal{G}(\overline{\pi}_1, \overline{\pi}_2)$. Our main result in the next section (Theorem 4) is to provide a positive answer to this. In fact, we will not only prove the existence of NE vectors but also provide a method to construct them.

We will end this section with the following theorem which is similar to Theorem 1-(b), that was used to obtain NEPPs. This theorem will enable us to construct PO-NEPPs.

Theorem 3: Given a PO policy pair $(\overline{\pi}^*_1, \overline{\pi}^*_2)$, construct the strategy vector pair $\{(f^*_\ell, g^*_\ell) : \ell \in \mathcal{L}\}$ as follows: $f^*_\ell(r_i) = \overline{\pi}^*_1(r_i, \ell)$ and $g^*_\ell(r_j) = \overline{\pi}^*_2(\ell, r_j)$ for all $i, j \in [n]$. Now, suppose for each ℓ, (f^*_ℓ, g^*_ℓ) is a NE vector for the game $\mathcal{G}(\overline{\pi}^*_1, \overline{\pi}^*_2)$ such that,

$$\min\left\{C^{(1)}_{\mathsf{s},g^*_\ell}(r_i), C^{(1)}_{\mathsf{c},g^*_\ell}\right\} = G^{(1)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(r_i, \ell), \text{ and} \tag{37}$$

$$\min\left\{C^{(2)}_{\mathsf{s},f^*_\ell}(r_j), C^{(2)}_{\mathsf{c},f^*_\ell}\right\} = G^{(2)}_{\overline{\pi}^*_1,\overline{\pi}^*_2}(\ell, r_j). \tag{38}$$

Then $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ is a PO-NEPP.

Proof: See Appendix E. ∎

Discussion: If $\{(f^*_\ell, g^*_\ell)\}$ happens to be a NE vector, then from Definition 7 it simply follows that the LHS of (37) (resp. (38)) is simply the cost incurred by $\mathscr{F}_1$ (resp. $\mathscr{F}_2$) for playing the action, $f^*_\ell(r_i)$ (resp. $g^*_\ell(r_j)$), suggested by its NE vector. Thus, (37) and (38) collective say that the cost-pair obtained by playing the NE vector (f^*_ℓ, g^*_ℓ) in the Bayesian game $\mathcal{G}(\overline{\pi}^*_1, \overline{\pi}^*_2)$, is equal to the cost-pair incurred by the PO policy pair $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ in the original POSG. Hence, this result could be thought of as the analogue of Theorem 1-(b) proved for the completely observable case.

Existence of a NE Vector: We will fix a PO policy pair $(\overline{\pi}^*_1, \overline{\pi}^*_2)$ that satisfies the inequalities in (33). In this section we will prove that there exists a NE vector for $\mathcal{G}(\overline{\pi}^*_1, \overline{\pi}^*_2)$. Before proceeding to the main theorem we need the following results (Lemma 4 and 5).

Lemma 4: For any $\ell \in \mathcal{L}$, the best response vector, f_ℓ, against any vector g_ℓ of $\mathscr{F}_2$ is a *threshold vector*, i.e., there exists an $\Phi_\ell \in \{0, 1, \cdots, n\}$ such that $f_\ell(r_i) = \mathsf{s}$ iff $i > \Phi_\ell$. We refer to Φ_ℓ as the *threshold of* f_ℓ. Similarly, if g_ℓ is the best response against any vector f_ℓ of $\mathscr{F}_1$, then g_ℓ is a threshold vector with threshold Ψ_ℓ.

Proof: Since $r_{i'} \le r_i$ whenever $i' \le i$, we can write $C^{(1)}_{\mathsf{s},g_\ell}(r_{i'}) \ge C^{(1)}_{\mathsf{c},g_\ell}(r_i)$ (see (35)). Then the proof follows by recalling Definition 6. ∎

Remark: The above lemma is possible primarily because of the independence assumption we had imposed at the beginning of Section VI. Suppose we had worked with the model where, given only r_i, $\mathscr{F}_1$'s belief about $\mathscr{F}_2$'s observation is simply the conditional p.m.f. $p_{R_1,R_2}(r_j|r_i)$, $j \in [n]$, then, as in (34), we can write the expression for the continuing probability as

$$\widetilde{g}_{\ell,r_i} = \sum_{j:g_\ell(r_j)=\mathsf{c}} p_{R_1,R_2}(r_j|r_i), \tag{39}$$

which is now a function of r_i. If we replace $\widetilde{g}_\ell$ in (35) by $\widetilde{g}_{\ell,r_i}$ it is not possible to conclude, $C^{(1)}_{s,g_\ell}(r_{i'}) \ge C^{(1)}_{s,g_\ell}(r_i)$ whenever $i' \le i$, as required for the proof of the above lemma.

The following is an immediate consequence of Lemma 4: if (f^*_ℓ, g^*_ℓ) is a NE vector then f^*_ℓ and g^*_ℓ are both threshold vectors. Thus, we can restrict our search for NE vectors over the class of all pairs of threshold vectors. Since a threshold vector f_ℓ can be equivalently represented by its threshold Φ_ℓ we can alternatively work with the thresholds. Thus $\Phi_\ell \in \mathcal{A}_0 := \{0, 1, \cdots, n\}$ represents the $n+1$ thresholds that $\mathscr{F}_1$ can use. $\Phi_\ell = 0$ (respectively, $\Phi_\ell = n$) represents the threshold vector which, when used by $\mathscr{F}_1$, stops (respectively, continues) for any value of r_i when the location is ℓ. Similarly, we will represent the $n+1$ thresholds that $\mathscr{F}_2$ can use by $\Psi_\ell \in \mathcal{A}_0$. We will write $\Phi_\ell = BR_1(\Psi_\ell)$ whenever their corresponding threshold vectors, f_ℓ and g_ℓ, respectively, are such that $f_\ell = BR_1(g_\ell)$. Similarly, we will write $\Psi_\ell = BR_2(\Phi_\ell)$ whenever $g_\ell = BR_2(f_\ell)$.

Lemma 5: (1) Let $\Psi_\ell, \Psi^o_\ell \in \mathcal{A}_0$ be two thresholds of $\mathscr{F}_2$ such that $\Psi_\ell < \Psi^o_\ell$, then the best response of $\mathscr{F}_1$ to these are ordered as, $BR_1(\Psi_\ell) \ge BR_1(\Psi^o_\ell)$. (2) Similarly, if $\Phi_\ell, \Phi^o_\ell \in \mathcal{A}_0$ are two thresholds of $\mathscr{F}_1$ such that $\Phi_\ell < \Phi^o_\ell$ then $BR_2(\Phi_\ell) \ge BR_2(\Phi^o_\ell)$.

Proof: See Appendix F. ∎

We are now ready to prove the following main theorem. We will present the complete proof here because the proof technique will be required in the next section to construct PO-NEPPs.

Theorem 4: For every $\ell \in \mathcal{L}$, there exists a NE vector (f^*_ℓ, g^*_ℓ) for the game $\mathcal{G}(\overline{\pi}^*_1, \overline{\pi}^*_2)$.

Proof: As mentioned earlier, a consequence of Lemma 4 is that it is sufficient to restrict our search for NE vectors within the class of all pairs of threshold vectors. Let $\mathcal{A}_0 := \{\Phi_\ell : 0 \le \Phi_\ell \le n\}$ denote the set of all $n+1$ thresholds of $\mathscr{F}_1$. Now, for $1 \le k \le n$, inductively define the sets $\mathcal{B}_k$ and $\mathcal{A}_k$ as follows: $\mathcal{B}_k = \left\{BR_2(\Phi_\ell) : \Phi_\ell \in \mathcal{A}_{k-1}\right\}$ and $\mathcal{A}_k = \left\{BR_1(\Psi_\ell) : \Psi_\ell \in \mathcal{B}_k\right\}$.

It is easy to check that through this *inductive process* we will finally end up with non-empty sets $\mathcal{A}_n$ and $\mathcal{B}_n$ such that

- for each $\Phi_\ell \in \mathcal{A}_n$ there exists a unique $\Psi_\ell \in \mathcal{B}_n$ such that $\Phi_\ell = BR_1(\Psi_\ell)$, and
- for each $\Psi_\ell \in \mathcal{B}_n$ there exists a unique $\Phi_\ell \in \mathcal{A}_n$ such that $\Psi_\ell = BR_2(\Phi_\ell)$.

Since best responses are unique, these would also mean that $|\mathcal{A}_n| = |\mathcal{B}_n|$.

Note that there is nothing special about this inductive process, in the sense that for any normal form game with two player, each of whose action set is $\mathcal{A}_0$, this inductive process will still yield sets $\mathcal{A}_n$ and $\mathcal{B}_n$ satisfying the above properties whenever the best responses are unique. However, it is possible that there exists no pair $(\Phi_\ell, \Psi_\ell) \in \mathcal{A}_n \times \mathcal{B}_n$ such that $\Phi_\ell = BR_1(\Psi_\ell)$ and $\Psi_\ell = BR_2(\Phi_\ell)$. For instance, $\mathcal{A}_n = \{\Phi_\ell, \Phi'_\ell\}$, $\mathcal{B}_n = \{\Psi_\ell, \Psi'_\ell\}$ and $BR_2(\Phi_\ell) = \Psi_\ell$ and $BR_2(\Phi'_\ell) = \Psi'_\ell$ while $BR_1(\Psi_\ell) = \Phi'_\ell$ and $BR_1(\Psi'_\ell) = \Phi_\ell$. This is precisely where Lemma 5 will be useful, due to which such a situation cannot arise in our case.

Now, arrange the $N = |\mathcal{A}_n|$ $(= |\mathcal{B}_n|)$ remaining thresholds in $\mathcal{A}_n$ and $\mathcal{B}_n$ as, $\Phi_{\ell,1} < \Phi_{\ell,2} < \cdots < \Phi_{\ell,N}$ and $\Psi_{\ell,1} < \Psi_{\ell,2} < \cdots < \Psi_{\ell,N}$, respectively. Then $\Phi_{\ell,1} = BR_1(\Psi_{\ell,N})$, since if not then using Lemma 5 we can write $\Phi_{\ell,1} < BR_1(\Psi_{\ell,N}) \leq BR_1(\Psi_{\ell,t})$ for every $t = 1, 2, \cdots, N$ contradicting the fact that $\Phi_{\ell,1}$ being in $\mathcal{A}_n$ has to be the best response for some $\Psi_{\ell,t} \in \mathcal{B}_n$. Similarly $\Psi_{\ell,N} = BR_2(\Phi_{\ell,1})$, otherwise again from Lemma 5 we obtain $\Psi_{\ell,N} > BR_2(\Phi_{\ell,1}) \geq BR_2(\Phi_{\ell,t})$ for every $t = 1, 2, \cdots, N$ leading to a contradiction that $\Psi_{\ell,N}$ is not the best response of any $\Phi_{\ell,t} \in \mathcal{A}_n$. Thus the threshold strategy pair (f^*_ℓ, g^*_ℓ) corresponding to the threshold pair $(\Phi_{\ell,1}, \Psi_{\ell,N})$ is a NE vector. By an inductive argument, it can be shown that all the threshold vector pairs corresponding to the threshold pairs $(\Psi_{\ell,t}, \Psi_{\ell,N-(t-1)})$, $t = 1, 2, \cdots, N$, are NE vectors. ■

D. PO-NEPP Construction from NE Vectors

Once we have obtained NE vectors (f^*_ℓ, g^*_ℓ), for each $\ell \in [m]$, The procedure for constructing PO-NEPP from NE vectors is along the same lines as the construction of NEPP from NE strategies (see Section V-C).

We begin with a pair of cost terms, $\overline{\mathbf{C}} = (\overline{C}^{(1)}, \overline{C}^{(2)})$, satisfying (33). Using the procedure in the proof of Theorem 4, we obtain, for each $\ell \in \mathcal{L}$, the NE vector $(f^\nabla_\ell, g^\nabla_\ell)$ corresponding to the threshold pair $(\Phi_{\ell,1}, \Psi_{\ell,N})$ ($\mathscr{F}_1$ using lowest threshold while $\mathscr{F}_2$ uses the highest). Then we define

$$\begin{aligned} G^{(1)}(r_i, \ell) &= \min\Big\{C^{(1)}_{\mathsf{s},g^\nabla_\ell}(r_i), C^{(1)}_{\mathsf{c},g^\nabla_\ell}\Big\} \\ G^{(2)}(\ell, r_j) &= \min\Big\{C^{(2)}_{\mathsf{s},f^\nabla_\ell}(r_j), C^{(2)}_{\mathsf{c},f^\nabla_\ell}\Big\}. \end{aligned}$$

Now recall the expressions for the costs $\overline{C}^{(1)}$ and $\overline{C}^{(2)}$ from (31) and (32). Compute the RHS of these expressions by replacing $G^{(1)}_{\pi^*_1,\pi^*_2}(\cdot)$ and $G^{(2)}_{\pi^*_1,\pi^*_2}(\cdot)$ by the functions $G^{(1)}(\cdot)$ and $G^{(2)}(\cdot)$, respectively. Denote the computed sums as $\overline{\mathcal{T}}_1(\overline{\mathbf{C}})$ and $\overline{\mathcal{T}}_2(\overline{\mathbf{C}})$, respectively. Suppose $\overline{\mathbf{C}}$ is such that $\overline{\mathbf{C}} = (\overline{\mathcal{T}}_1(\overline{\mathbf{C}}), \overline{\mathcal{T}}_2(\overline{\mathbf{C}}))$ (we inductively iterate to obtain such a $\overline{\mathbf{C}}$) then using Theorem 3 we can construct the PO-NEPP, $(\overline{\pi}^\nabla_1, \overline{\pi}^\nabla_2)$ using $(f^\nabla_\ell, g^\nabla_\ell)$ as follows: for each $i, j \in [n]$ and $\ell \in \mathcal{L}$, $\overline{\pi}^\nabla_1(r_i, \ell) = f^\nabla_\ell(r_i)$ and $\overline{\pi}^\nabla_2(\ell, r_j) = g^\nabla_\ell(r_j)$.

Finally, since the threshold vector $(f^\Delta_\ell, g^\Delta_\ell)$ corresponding to the threshold pair $(\Phi_{\ell,N}, \Psi_{\ell,1})$ ($\mathscr{F}_1$ using highest threshold while $\mathscr{F}_2$ uses the lowest) is also a NE vector, one can similarly construct the PO-NEPP, $(\overline{\pi}^\Delta_1, \overline{\pi}^\Delta_2)$, using $(f^\Delta_\ell, g^\Delta_\ell)$.

VII. Cooperative Case

It will be interesting to benchmark the best performance that can be achieved if both forwarders would *cooperate* with each other. In this section, we will describe this case and construct a *Pareto optimal* performance curve.

We will assume the completely observable case. The definition of a policy pair (π_1, π_2) and the costs $J^{(1)}_{\pi_1,\pi_2}(x)$ and $J^{(2)}_{\pi_1,\pi_2}(x)$ will remain as in Section V. However, here our objective is instead to optimize a linear combination of the two costs. Formally, let $\gamma \in (0,1)$, then the problem we are interested in is,

$$\text{Minimize}_{(\pi_1,\pi_2)} \Big(\gamma J^{(1)}_{\pi_1,\pi_2}(x) + (1-\gamma) J^{(2)}_{\pi_1,\pi_2}(x)\Big). \tag{40}$$

Let $(\pi^\gamma_1, \pi^\gamma_2)$ denote the policy pair which is optimal for the above problem. Then, using (18) and (19), it is easy to show that $(\pi^\gamma_1, \pi^\gamma_2)$ is also optimal for

$$\text{Minimize}_{(\pi_1,\pi_2)} \Big(\gamma C^{(1)}_{\pi_1,\pi_2} + (1-\gamma) C^{(2)}_{\pi_1,\pi_2}(x)\Big). \tag{41}$$

We have the following lemma.

Lemma 6: The policy pair $(\pi^\gamma_1, \pi^\gamma_2)$ is *Pareto optimal*, i.e., for any other policy (π_1, π_2),

(1) if $C^{(1)}_{\pi_1,\pi_2} < C^{(1)}_{\pi^\gamma_1,\pi^\gamma_2}$ then $C^{(2)}_{\pi^\gamma_1,\pi^\gamma_2} < C^{(2)}_{\pi_1,\pi_2}$, and
(2) if $C^{(2)}_{\pi_1,\pi_2} < C^{(2)}_{\pi^\gamma_1,\pi^\gamma_2}$ then $C^{(1)}_{\pi^\gamma_1,\pi^\gamma_2} < C^{(1)}_{\pi_1,\pi_2}$.

Proof: Available in Appendix G. ■

Thus, by varying $\gamma \in (0,1)$, we obtain a Pareto optimal boundary whose points are $(C^{(1)}_{\pi^\gamma_1,\pi^\gamma_2}, C^{(2)}_{\pi^\gamma_1,\pi^\gamma_2})$. Details on how to obtain $(\pi^\gamma_1, \pi^\gamma_2)$ is available in Appendix G.

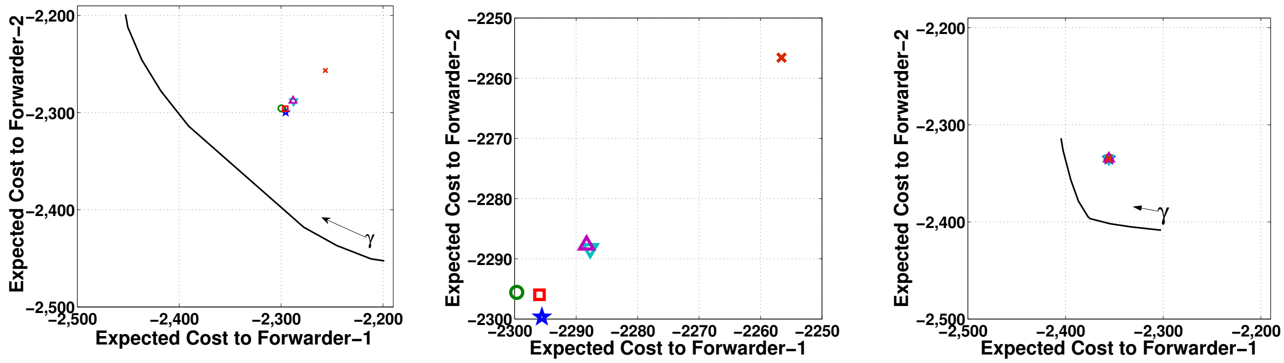
VIII. Numerical and Simulation Results for the Geographical Forwarding Example

A. One-Hop Study

The one-hop study can be more general, requiring only a joint p.m.f. $p_{i,j}$, a location p.m.f. q_ℓ, and conditional p.m.f.s $p^{(1)}_{i|\ell}$ and $p^{(2)}_{j|\ell}$ (for all i, j and ℓ). However, to illustrate the practicality of our study, we will study the geographical forwarding example described in Section III.

Recall the packet forwarding scenario illustrated in Fig. 1. We will fix the locations of $\mathscr{F}_1$ and $\mathscr{F}_2$ to be $v_1 = [0, \frac{\theta}{2}]$ and $v_2 = [0, -\frac{\theta}{2}]$, respectively. Thus, the distance of separation between the two forwarders is θ meters (m); we will vary θ and study the performance of the various policies. The range of each forwarder is $d = 80$ m. The combined forwarding region is discretized into a uniform grid where the distance between the neighboring points is 5 m. Finally, the sink node is placed at $v_0 = [1000, 0]$.

Next, recall the power and reward expressions from (5) and (6), respectively. We have fixed $d_{ref} = 5$ m, $\xi = 2.5$, and $a = 0.5$. For ΓN_0, which is referred to as the *receiver sensitivity*, we use a value of 10^{-9} milliWatts (mW) (equivalently -90 dBm) specified for the Crossbow TelosB wireless mote [43]. The maximum transmit power available at a node is $P_{max} = 1$ mW (equivalently 0 dBm; again from



(a) (b) (c)

Fig. 3. Performance of the various NEPPs and PO-NEPPs are depicted as points in $\Re^2$ where the first (second) coordinate is the expected cost incurred by $\mathscr{F}_1$ ($\mathscr{F}_2$). Fig. (a) corresponds to the case when the distance of separation $\theta = 0$ m. A portion of Fig. (a) is enlarged and shown in Fig. (b). Fig. (c) corresponds to $\theta = 10$ m.

the Crossbow TelosB data sheet). We allow for four different channel gain values: 0.4×10^{-3}, 0.6×10^{-3}, 0.8×10^{-3}, and 1×10^{-3}, each occurring with equal probability. Finally, we fix $\eta_1 = \eta_2 = 100$ (recall that η_ρ is the parameter used to trade-off between delay and reward (see (2)), $\nu_1 = 1 - \nu_2 = 0.5$ (ν_ρ is the probability that F_ρ will win the contention), and the mean inter-wake-up time $\tau = 10$ milliseconds (ms).

We first set $\theta = 0$ m (recall that θ is the distance between the two forwarders) and, in Fig. 3(a), depict the performance of various NEPPs and PO-NEPPs as pair of costs $\mathbf{C} = (C^{(1)}, C^{(2)})$ where $C^{(\rho)}$ is the cost incurred by F_ρ starting from time 0 if the particular NEPP or PO-NEPP is used. Also shown in Fig. 3(a) is the performance of a *simple policy* (the point marked $\times$; to be describe next) along with the Pareto optimal boundary (the solid curve). Since, from Fig. 3(a) it is not easy to distinguish between the various points, we show a section of Fig. 3(a) as Fig. 3(b). Fig. 3(c) corresponds to $\theta = 10$ m.

Various Policy Pairs: The description of various points seen in Fig. 3 is as follows (we will use $\mathbf{C}_{symbol}$ to denote the cost pair corresponding to the policy $symbol$):

- $\bigstar$,$\bigcirc$,and $\square$: performances of the NEPPs that uses the NE strategies $(\mathtt{s}, \mathtt{c})$, $(\mathtt{c}, \mathtt{s})$, and the mixed strategy (Γ_1, Γ_2), respectively, whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_\bigstar)$, $\mathcal{R}_4(\mathbf{C}_\bigcirc)$,and $\mathcal{R}_4(\mathbf{C}_\square)$, respectively (recall Fig. 2).
- ∇ and Δ: performances of the PO-NEPPs that are constructed by choosing, for each $\ell \in \mathcal{L}$, the thresholds $(\Phi_{\ell,1}, \Psi_{\ell,N})$ and $(\Phi_{\ell,N}, \Psi_{\ell,1})$, respectively (recall the proof of Theorem 4).
- $\times$: performance of a simple policy where each forwarder $\mathscr{F}_\rho$ $(\rho = 1, 2)$ chooses $\mathtt{s}$ if and only if its reward value $r_\rho \geq \alpha^{(\rho)}$. Such a policy is optimal whenever $\mathscr{F}_\rho$ is alone in the system (recall (12) and (15)). Thus, using the simple policy each forwarder behaves as if the competing forwarder is not present.
- solid curve: Pareto optimal boundary obtained by $(\pi_1^\gamma, \pi_2^\gamma)$, $\gamma \in (0, 1)$; recall Section VII.

Observations: From Fig. 3(b) we see that operating at NEPP $\bigstar$ is most favorable for $\mathscr{F}_2$ since $C^{(2)}_\bigstar$ is less than the cost to $\mathscr{F}_2$ at the other two NEPPs, $C^{(2)}_\bigcirc$ and $C^{(2)}_\square$. This is because whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_\bigstar)$ the joint-action $(\mathtt{s}, \mathtt{c})$ played by $\bigstar$ fetches the least cost (of $D^{(2)}$) possibly by any strategy. In contrast, $\mathscr{F}_1$ incurs highest cost (of $-\eta_1 r_i$) possible because of which NEPP $\bigstar$ is least favorable for $\mathscr{F}_1$. For a similar reason, operating at NEPP $\bigcirc$ is most favorable for $\mathscr{F}_1$ while being least favorable for $\mathscr{F}_2$. The NEPP $\square$ which chooses the mixed strategy (Γ_1, Γ_2) whenever $(r_i, r_j) \in \mathcal{R}_4(\mathbf{C}_\square)$ helps to achieve a fairer cost to both players, however the performance at $\square$ is slightly farther from the Pareto boundary when compared with the other two NEPPs.

The performance at the PO-NEPPs, ∇ and Δ, is worse than at the NEPPs thus exhibiting the loss in performance due to partial information. The PO-NEPP ∇ which uses the NE vector corresponding to the lowest-highest best response pair, $(\Phi_{\ell,1}, \Psi_{\ell,N})$ (for each $\ell \in \mathcal{L}$), provides lower cost to $\mathscr{F}_2$ than the PO-NEPP Δ. This is because, $\mathscr{F}_1$ using a lower threshold will essentially choose an initial relay, thus leaving $\mathscr{F}_2$ alone in the system which can now accrue a better cost. For a similar reason, operating at Δ leads to $\mathscr{F}_1$ achieving a lower cost. Finally, the simple policy $\times$ has the worst performance in comparison with all other points, suggesting that it may not be wise to be operating using this policy pair. However, as we increase the value of θ the performance of the simple policy improves, and interestingly for $\theta = 10$ m (which is only 12.5% of the forwarders' range of 80 m) we observe that the various points are practically indistinguishable (note that the magnitude of the scales in plots Fig. 3(a) and 3(c) is the same). We have observed similar trend when $\eta_1 = \eta_2$ and a are set to different values.

Key Insight: Thus, based on our numerical work we draw the following key insight: even for a small distance of separation between the forwarders, using the simple policy pair (where each forwarder behaves as if it is alone in the system) yields little (or, practically, no) loss in performance when compared with the performance of an NEPP or a PO-NEPP; however the performance degradation of the simple policy is significant whenever the forwarders are very close to each other. These observations are for the case where there are two forwarders. However, we expect a similar behavior for the simple policy even if there are more than two forwarders, i.e., we believe that the simple policy performs well if the competing forwarders are moderately separated.

B. End-to-End Study

Finally, in this section we use simulation to provide an evaluation of the end-to-end performance of local forwarding. The competitive forwarding policies (i.e., NEPP and PO-NEPP) are difficult to implement since their structure has to be evaluated for each forwarding instance along the path of a packet. However, based on our observations in the previous section, we study the performance of the simple policy pair. In our prior work we have already studied the simple policy's performance (see [4, Fig. 8] where the simple policy is referred to as SF), but there the setting was that of the *lone packet model* where a single alarm packet is generated which is then routed to the sink. Here, we will generalize the lone packet setting by generating multiple packets simultaneously across the network so that a packet, along its route, might have to compete with other packets in its vicinity before reaching the sink.

We first form a network by randomly placing 1000 nodes in a square region of area 1 Km^2. A source node is placed at $[0, 1000]$ followed by a sink node at the diagonally opposite corner $[1000, 0]$. Each node is allowed to asynchronously and periodically sleep-wake cycle with period $T = 100$ ms, i.e., each node i wakes up and stays ON for a small duration (which we neglect, given the other time scales) at the periodic instants $T_i + kT$, $k \geq 1$ where $\{T_i\}$ are i.i.d. uniform on $[0, T]$ (recall the discussion on the sleep-wake process from Section II).

Each node i, assuming an inter-wake-up time of $1/N_i$ (where N_i is the average number of nodes in the forwarding region of node i), obtains α_i which is the threshold (on reward) required to implement the simple policy by node i. The values of all the other parameters required to compute the threshold, e.g., P_{max}, ξ, etc., remain the same as in our one-hop study. If there is no relay whose reward value is more than α_i (node i will know of this after waiting for one entire duty-cycling period T), node i, at time T, will simply forward the packet to the relay with the maximum reward (thus, as relays wake-up the best relay so far, is asked to wait).

The source node generates an alarm packet at time 0. We introduce competition by generating additional packets at randomly chosen nodes, randomly in time at the points of a Poisson process of rate λ. All the packets are destined for the same sink. While forwarding, if a relay is chosen simultaneously by more than one forwarder, then randomly one of them will win the contention and gets the relay to forward its packet to. We are interested in studying, as a function of λ, the performance obtained (in terms of end-to-end delay and the total power expended) in routing the source's packet.

In Fig. 4 we have plotted, for different values of λ, the mean end-to-end delay vs. the mean end-to-end power (averaged over packets from the source located at $(0, 1000)$). These curves are obtained by varying η, the parameter used to trade-off between delay and reward in the local problem. Each data point in Fig. 4 is the average of the respective quantities over 100 alarm packets generated by the source node. Also shown in the figure is the performance curve corresponding to the "lone packet case" where no additional packets are generated.

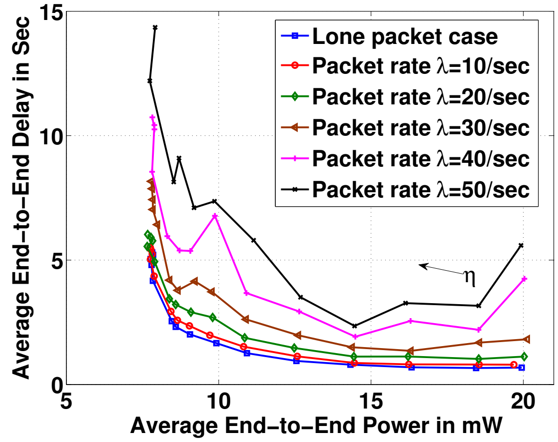


Fig. 4. End-to-end performance (average power vs. average delay) of the simple policy as the competing packet rate λ in the network is increased.

Hence the lone packet curve is analogous to the SF policy's performance curves in [4, Fig. 8].

Observe that, as we increase λ we obtain a *degradation in performance*, i.e., increased delay and power compared with the lone packet case. This is because, as λ increases, since there are more packets in the network, there is a larger probability that a forwarding node carrying the source's packet has to compete with other packets in the process of acquiring a relay. Also, as λ increases, at these instances of competition, the competing nodes tend to be closer together. From the observations in the previous section, we can conclude that as λ increases the performance of the simple policy will progressively degrade. However, the performance degradation is only marginal when the packet rate $\lambda \leq 20$ packets/sec while being moderate for $\lambda = 30$ packets/sec, thus supporting the usage of the simple policy for these packet rates. For higher values of λ (e.g., $\lambda = 40$ packets/sec and beyond) the performance degradation is significant and hence there could be a benefit in using NEPPs to forward packets for these rates.

Finally, we have only presented simulation results for the simple policy, since implementing NEPPs or PO-NEPPs for end-to-end routing has the following difficulties: (1) for a given pair of neighboring nodes, obtaining NEPPs will require fixed point iterations, (2) NEPPs are node pair dependent, so that all possible neighboring node pairs are required to compute the corresponding NEPPs, since during actual forwarding a node may be competing with any of its neighbors. Thus, there is a large complexity involved in implementing NEPPs. In contrast, the simple policy (being a single threshold based) is easy to implement. Moreover, for realistic parameter values corresponding to TelosB wireless mote, we have seen that the performance of simple policy is good (in comparison with the lone packet case) for packet rates $\lambda \leq 30$ packets/sec.

IX. Conclusion

We studied the problem of competitive relay selection when two forwarders compete for a next-hop relay (or some resource in general). We first considered the model where complete information is available to both the forwarders. We formulated the problem as a stochastic game and proceeded to obtain solution in terms of Nash equilibrium policy pairs (NEPPs).

We were able to provide insight into the structure of NEPPs, which was primarily possible because of our following key result (Lemma 2): "cost of continuing alone" is less than the "cost of continuing along with a competing forwarder". We next studied a partially observable case for which we constructed a Bayesian game which is effective played at each stage. For this Bayesian game, we proved the existence of a Nash equilibrium strategy within the class of (pure) threshold vectors (Theorem 4). The proof method of this result enabled us to construct NEPPs for the partial case. For the geographical forwarding example, through numerical experiments we observed that, even for moderate separation between the two forwarders, the performance of our simple policy is as good as the performance of any other NEPP/PO-NEPP. In the context of end-to-end forwarding, through simulations we established (for the considered setting) that for packet rates less than 30 packets/second, the performance of the simple policy is good compared with the lone packet case.

References

[1] Y. Yao, S. Ngoga, D. Erman, and A. Popescu, "Competition-Based Channel Selection for Cognitive Radio Networks," in *IEEE Wireless Communications and Networking Conference*, April 2012.

[2] D. Niyato and E. Hossain, "Competitive Spectrum Sharing in Cognitive Radio Networks: A Dynamic Game Approach," *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2651–2660, July 2008.

[3] J. Kim, X. Lin, and N. Shroff, "Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 2, pp. 484 –497, April 2011.

[4] K. P. Naveen and A. Kumar, "Relay Selection for Geographical Forwarding in Sleep-Wake Cycling Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 3, pp. 475–488, 2013.

[5] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Ad Hoc Networks*, vol. 3, pp. 325–349, 2005.

[6] M. Mauve, J. Widmer, and H. Hartenstein, "A Survey on Position-Based Routing in Mobile Ad-Hoc Networks," *IEEE Network*, vol. 15, pp. 30–39, 2001.

[7] E. Cinlar, *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

[8] A. Kumar, D. Manjunath, and J. Kuri, *Wireless Networking*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2008.

[9] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY, USA: Cambridge University Press, 2005.

[10] K. P. Naveen and A. Kumar, "Relay Selection with Channel Probing in Sleep-Wake Cycling Wireless Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 11, no. 3, pp. 52:1–52:38, May 2015.

[11] H. Takagi and L. Kleinrock, "Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals," *IEEE Transactions on Communications [legacy, pre - 1988]*, vol. 32, no. 3, pp. 246–257, 1984.

[12] T. C. Hou and V. Li, "Transmission Range Control in Multihop Packet Radio Networks," *IEEE Transactions on Communications*, vol. 34, no. 1, pp. 38–44, 1986.

[13] R. Nelson and L. Kleinrock, "The Spatial Capacity of a Slotted ALOHA Multihop Packet Radio Network with Capture," *IEEE Transactions on Communications*, vol. 32, no. 6, pp. 684–694, 1984.

[14] M. Zorzi and R. R. Rao, "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance," *IEEE Transactions on Mobile Computing*, vol. 2, pp. 337–348, 2003.

[15] S. Liu, K. W. Fan, and P. Sinha, "CMAC: An Energy Efficient MAC Layer Protocol using Convergent Packet Forwarding for Wireless Sensor Networks," in *SECON '07, 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, June 2007.

[16] P. Chaporkar and A. Proutiere, "Optimal Joint Probing and Transmission Strategy for Maximizing Throughput in Wireless Systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1546 –1555, October 2008.

[17] N. B. Chang and M. Liu, "Optimal Channel Probing and Transmission Scheduling for Opportunistic Spectrum Access," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, 2007.

[18] Freeman, "The secretary problem and its extensions: A review," *International Statistical Review*, 1983.

[19] R. N. Bradt, S. M. Johnson, and S. Karlin, "On Sequential Designs for Maximizing the Sum of n Observations," *The Annals of Mathematical Statistics*, vol. 27, no. 4, pp. 1060–1074, 12 1956.

[20] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I.* Athena Scientific, 2005.

[21] S. Karlin, *Stochastic Models and Optimal Policy for Selling an Asset, Studies in Applied Probability and Management Science / edited by Kenneth J. Arrow, Samuel Karlin, Herbert Scarf* . Stanford University Press, Stanford, Calif, 1962.

[22] I. David and O. Levi, "A New Algorithm for the Multi-item Exponentially Discounted Optimal Selection Problem," *European Journal of Operational Research*, vol. 153, no. 3, pp. 782 – 789, 2004.

[23] B. K. Kang, "Optimal Stopping Problem with Double Reservation Value Property," *European Journal of Operational Research*, vol. 165, no. 3, pp. 765 – 785, 2005.

[24] S. C. Albright, "A Bayesian Approach to a Generalized House Selling Problem," *Management Science*, vol. 24, no. 4, pp. 432–440, 1977.

[25] D. B. Rosenfield, R. D. Shapiro, and D. A. Butler, "Optimal Strategies for Selling an Asset," *Management Science*, vol. 29, no. 9, pp. 1051–1061, 1983.

[26] J.-I. Nakagami, "A Two-Person Noncooperative Game for Assets Selling Problem: Independent Case," *Computers and Mathematics with Applications*, vol. 37, pp. 207 – 212, 1999.

[27] N. Immorlica, R. Kleinberg, and M. Mahdian, "Secretary Problems with Competing Employers," in *Internet and Network Economics*, ser. Lecture Notes in Computer Science, P. Spirakis, M. Mavronicolas, and S. Kontogiannis, Eds. Springer Berlin Heidelberg, 2006, vol. 4286, pp. 389–400.

[28] D. Ramsey and K. Szajowski, "Bilateral Approach to the Secretary Problem," in *Advances in Dynamic Games*, ser. Annals of the International Society of Dynamic Games, A. Nowak and K. Szajowski, Eds. Birkhuser Boston, 2005, vol. 7, pp. 271–284.

[29] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. New York, NY, USA: Springer-Verlag New York, Inc., 1996.

[30] A. Fink, "Equilibrium in a Stochastic n-Person Game," *Hiroshima Mathematical Journal*, vol. 28, no. 1, pp. 89–93, 1964.

[31] F. Thuijsman and O. J. Vrieze, "Total Reward Stochastic Games and Sensitive Average Reward Strategies," *J. Optim. Theory Appl.*, vol. 98, no. 1, pp. 175–196, July 1998.

[32] E. Altman, "Non Zero-Sum Stochastic Games in Admission, service and Routing Control in Queueing Systems," *Queueing Systems*, vol. 23, pp. 259–279, 1996.

[33] E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, and D. S. Menasche, "Constrained Cost-Coupled Stochastic Games with Independent State Processes," *Operations Research Letters*, vol. 36, no. 2, pp. 160 – 164, 2008.

[34] T. Raghavan and J. Filar, "Algorithms for Stochastic Games - A Survey," *Zeitschrift fur Operations Research*, vol. 35, pp. 437–472, 1991.

[35] L. S. Shapley, "Stochastic Games," *Proceedings of the National Academy of Sciences*, vol. 39, no. 10, pp. 1095–1100, 1953.

[36] A. Nowak and K. Szajowski, *Advances in Dynamic Games: Applications to Economics, Finance, Optimization, and Stochastic Control*, ser. Annals of the International Society of Dynamic Games. Birkhauser Boston, 2007.

[37] E. Solan and R. V. Vohra, "Correlated Equilibrium in Quitting Games," *Mathematics of Operations Research*, vol. 26, no. 3, pp. 601–610, 2001.

[38] V. Pata, "Fixed Point Theorems and Applications," 2014. [Online]. Available: www.mate.polimi.it/viste/pagina_personale/pp/121/FP.pdf

[39] D. P. Bertsekas and J. N. Tsitsiklis, "An Analysis of Stochastic Shortest Path Problems," *Mathematics of Operations Research*, vol. 16, pp. 580–595, 1991.

[40] E. A. Hansen, D. S. Bernstein, and S. Zilberstein, "Dynamic Programming for Partially Observable Stochastic Games," in *Proceedings of the 19th national conference on Artifical intelligence*, ser. AAAI'04. AAAI Press, 2004, pp. 709–715.

[41] M. J. Osborne, *An Introduction to Game Theory*. Oxford University Press, USA, August 2003.

[42] L. MacDermed, C. Isbell, and L. Weiss, "Markov Games of Incomplete Information for Multi-Agent Reinforcement Learning," in *AAAI Workshops*, 2011.

[43] Crossbow, "TelosB Mote Platform," 2014. [Online]. Available: www.willow.co.uk/TelosB_Datasheet.pdf

APPENDIX A
PROOF OF LEMMA 1

For convenience, here in the appendix we will recall the respective Lemma/Theorem statement before providing its proof.

Lemma 1: $\alpha^{(1)}$ is the unique fixed point of $\beta^{(1)}(x)$ ($x \in (-\infty, r_n]$) in (13).

Proof: Let us recall the expression of $\beta^{(1)}(x)$:

$$\beta^{(1)}(x) = \mathbb{E}\Big[\max\{x, R_1\}\Big] - \frac{\tau}{\eta_1},$$

where the expection is w.r.t. the p.m.f. $p^{(1)}$ of R_1 (recall that R_1 takes values from the set $\{r_1, r_2, \cdots, r_n\}$).

Let $m = \max\{i \in [n] : p_i^{(1)} > 0\}$. For $x > r_m$, note that $\beta^{(1)}(x) = x - \frac{\tau}{\eta_1} < x$. Hence a fixed point, if any, should lie within $(-\infty, r_m]$. Let us restrict $\beta^{(1)}(\cdot)$ to the domain $(-\infty, r_m]$. Then, since $\beta^{(1)}(x) \le r_m$ for any $x \in (-\infty, r_m]$, we have $\beta^{(1)} : (-\infty, r_m] \to (-\infty, r_m]$. We can now proceed to show that $\beta^{(1)}(x)$ restricted to $x \in (-\infty, r_m]$ is a *contraction mapping*, i.e., for any $x, x' \in (-\infty, r_m]$, we need to show that

$$\| \beta^{(1)}(x) - \beta^{(1)}(x') \| \le \kappa \| x - x' \| \tag{42}$$

for some $\kappa < 1$. Without loss of generality let $x > x'$. Then,

$$\begin{aligned}
\| \beta^{(1)}(x) - \beta^{(1)}(x') \| &= \beta^{(1)}(x) - \beta^{(1)}(x') \\
&= \mathbb{E}\Big[\max\{x, R_1\}\Big] - \mathbb{E}\Big[\max\{x', R_1\}\Big] \\
&= \sum_{i=1}^{n} p_i^{(1)} \Big(\max\{x, r_i\} - \max\{x', r_i\} \Big) \\
&\overset{*}{=} \sum_{i=1}^{m} p_i^{(1)} \Big(\max\{x, r_i\} - \max\{x', r_i\} \Big) \\
&\overset{o}{=} \sum_{i=1}^{m-1} p_i^{(1)} \Big(\max\{x, r_i\} - \max\{x', r_i\} \Big) \\
&\overset{\dagger}{\le} \sum_{i=1}^{m-1} p_i^{(1)} \Big(x - x' \Big) \\
&= (1 - p_m^{(1)}) \| x - x' \| .
\end{aligned}$$

In the above derivation, $*$ is because $p_i^{(1)} = 0$ for $i > m$ (recall the definition of m); o is because, since $x, x' \le r_m$, we have $\Big(\max\{x, r_m\} - \max\{x', r_m\} \Big) = 0$; to obtain $\dagger$ note that, $\Big(\max\{x, r_i\} - \max\{x', r_i\} \Big) \le (x - x')$ for any r_i. Thus, $\beta^{(1)}(x)$, $x \in (-\infty, r_m]$ is a contraction mapping (recall 42) with $\kappa = (1 - p_m^{(1)}) < 1$ (since $p_m^{(1)} > 0$ from definition). Hence from the Banach fixed point theorem [38] it follows that there exists a unique fixed point $\alpha^* \in (-\infty, r_m]$, i.e., α^* satisfies $\alpha^* = \beta^{(1)}(\alpha^*)$.

Now, suppose we can show that

$$J^{(1)}_{\pi_1^*, \pi_2^*}(r_i, \boldsymbol{t}) = \min\Big\{ -\eta_1 r_i, -\eta_1 \alpha^* \Big\} \tag{43}$$

then, recalling the expression for $D^{(1)}$ from (11), we obtain

$$\begin{aligned}
\alpha^{(1)} &= \frac{D^{(1)}}{-\eta_1} \\
&= -\frac{\tau}{\eta_1} - \frac{1}{\eta_1} \sum_i p_i^{(1)} J^{(1)}_{\pi_1^*, \pi_2^*}(r_i, \boldsymbol{t}) \\
&= \mathbb{E}\Big[\max\{\alpha^*, R_1\}\Big] - \frac{\tau}{\eta_1}. \\
&= \beta^{(1)}(\alpha^*) \\
&= \alpha^*.
\end{aligned}$$

Thus, $\alpha^{(1)}$ is the unique fixed point of $\beta^{(1)}(\cdot)$.

To show (43), we proceed as follows. Let $J_0(r_i) = 0$ for all r_i, and for $k \ge 1$ define $J_k(r_i)$ inductively as

$$J_k(r_i) = \min\Big\{ -\eta_1 r_i, \tau + \mathbb{E}\Big[J_{k-1}(R_1)\Big]\Big\}. \tag{44}$$

Since our problem with one player is equivalent to the optimal stopping problem studied in [39], the above iterations converge to the optimal cost, i.e., $\lim_{k\to\infty} J_k(r_i) = J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t})$. Now, defining $\alpha_1 = -\frac{\tau}{\eta_1}$, $J_1(r_i)$ can be written as $J_1(r_i) = \min\{-\eta_1 r_i, -\eta_1\alpha_1\}$. Proceeding further we can write,

$$\begin{aligned} J_2(r_i) &= \min\Big\{ -\eta_1 r_i, \tau + \mathbb{E}\Big[J_1(R_1)\Big]\Big\} \\ &= \min\Big\{ -\eta_1 r_i, \tau + \mathbb{E}\Big[\min\{-\eta_1\alpha_1, -\eta_1 R\}\Big]\Big\} \\ &= \min\Big\{ -\eta_1 r_i, -\eta_1\beta^{(1)}(\alpha_1)\Big\} \\ &= \min\Big\{ -\eta_1 r_i, -\eta_1\alpha_2\Big\} \end{aligned}$$

where $\alpha_2 = \beta^{(1)}(\alpha_1)$. Similarly it can be shown that, if $J_{k-1}(r_i) = \min\Big\{ -\eta_1 r_i, -\eta_1\alpha_{k-1}\Big\}$, then

$$J_k(r_i) = \min\Big\{ -\eta_1 r_i, -\eta_1\alpha_k\Big\} \tag{45}$$

where $\alpha_k = \beta^{(1)}(\alpha_{k-1})$. Thus $\alpha_k \to \alpha^*$. Finally, in the above expression taking the limit as $k\to\infty$ on both sides, and using $J_k(r_i) \to J^{(1)}_{\pi_1^*,\pi_2^*}(r_i,\boldsymbol{t})$ and $\alpha_k \to \alpha^*$, we obtain the desired result. ■

APPENDIX B
PROOF OF THEOREM 1

Theorem 1: Given a policy pair, (π_1^*, π_2^*), construct the static game given in Table 5. Then the following statements are

	c	s
c	$C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*}$	$D^{(1)}, -\eta_2 r_j$
s	$-\eta_1 r_i, D^{(2)}$	$E^{(1)}(r_i), E^{(2)}(r_j)$

TABLE 5
STATIC STAGE GAME.

equivalent:

(a) (π_1^*, π_2^*) is an NEPP.
(b) For any $x = (r_i, r_j)$, $(\pi_1^*(x), \pi_2^*(x))$ is a *Nash equilibrium (NE) strategy* for the game in Table 5. Further, the expected cost pair at this NE strategy is, $(J^{(1)}_{\pi_1^*,\pi_2^*}(x), J^{(2)}_{\pi_1^*,\pi_2^*}(x))$.

Proof of (a) $\Longrightarrow$ (b): Suppose (a) is true, i.e., (π_1^*, π_2^*) is an NEPP. Then, π_1^* is the best response policy of $\mathscr{F}_1$ against the policy π_2^* of $\mathscr{F}_2$. Hence π_1^* is optimal for the MDP problem, denoted $MDP_1(\pi_2^*)$, which is obtained by fixing the policy π_2^* of $\mathscr{F}_2$ (note that $MDP_1(\pi_2^*)$ is a time homogeneous MDP since π_2^* is stationary; recall Definition 1). Since (1) the states of the form $(\boldsymbol{t}, r_j)$ are absorbing and cost free for $\mathscr{F}_1$, and (2) the policy of $\mathscr{F}_1$ which never stops incurs infinite cost to $\mathscr{F}_1$, it follows that $MDP_1(\pi_2^*)$ is an optimal stopping problem [39]. Hence, $J^{(1)}_{\pi_1^*,\pi_2^*}(x)$, $x = (r_i, r_j)$ satisfies the following Bellman equation,

$$\begin{aligned} J^{(1)}_{\pi_1^*,\pi_2^*}(x) &= \min\Big\{C_{\mathtt{s}}(x), C_{\mathtt{c}}(x)\Big\} \\ &= \min\Big\{\pi_2^*(x,\mathtt{c})(-\eta_1 r_i) + \pi_2^*(x,\mathtt{s})E^{(1)}(r_i), \\ &\qquad \pi_2^*(x,\mathtt{c})C^{(1)}_{\pi_1^*,\pi_2^*} + \pi_2^*(x,\mathtt{s})D^{(1)}\Big\}, \end{aligned} \tag{46}$$

where $\pi_2^*(x,\mathtt{c})$ (resp. $\pi_2^*(x,\mathtt{s})$) is the probability that $\mathscr{F}_2$ will choose action c (resp. s) when the state is x. The two terms in the min-expression above (denoted $C_{\mathtt{s}}(x)$ and $C_{\mathtt{c}}(x)$) are the expected cost to $\mathscr{F}_1$ for taking actions s and c, respectively. Note that these costs are exactly the expected cost incurred by $\mathscr{F}_1$, for playing actions s and c, respectively, in the static game in Table 4, when the strategy of $\mathscr{F}_2$ is $\pi_2^*(x)$. Now π_1^*, being optimal for $MDP_1(\pi_2^*)$, chooses action $\pi_1^*(x) \in \Delta(\{\mathtt{s},\mathtt{c}\})$ whichever gives a minimum cost or can randomize between the two if both the costs are equal. Hence, it follows from the structure of (46) that $\pi_1^*(x)$ is the best response against $\pi_2^*(x)$ for the game in Table 4. Further the cost to $\mathscr{F}_1$ for playing $\pi_1^*(x)$, from Table 4, is $\min\{C_{\mathtt{s}}(x), C_{\mathtt{c}}(x)\} = J^{(1)}_{\pi_1^*,\pi_2^*}(x)$.

Similarly, by writing the Bellman equation corresponding to the $MDP_2(\pi_1^*)$ problem (which is obtained by fixing the policy π_1^* of $\mathscr{F}_1$), we can conclude that $\pi_2^*(x)$ is the best response against $\pi_1^*(x)$ for the game in Table 4, with the cost to player $\mathscr{F}_2$ being $J^{(2)}_{\pi_1^*,\pi_2^*}(x)$.

Proof of (b) $\Longrightarrow$ (a): Given that the policy (π_1^*, π_2^*) satisfies the condition in (b), let π_1 be any policy of $\mathscr{F}_1$. Then, for any $x = (r_i, r_j)$, since $(\pi_1^*(x), \pi_2^*(x))$ is a NE strategy for the game in Table 4 with cost to $\mathscr{F}_1$ at equilibrium being $J^{(1)}_{\pi_1^*,\pi_2^*}(x)$, we can write

$$\begin{aligned} J^{(1)}_{\pi_1^*,\pi_2^*}(x) \quad &\leq \quad \pi_1(x,\mathsf{c})\Big(\pi_2^*(x,\mathsf{c})C^{(1)}_{\pi_1^*,\pi_2^*} + \pi_2^*(x,\mathsf{s})D^{(1)}\Big) + \\ &\qquad \pi_1(x,\mathsf{s})\Big(\pi_2^*(x,\mathsf{c})(-\eta_1 r_i) + \pi_2^*(x,\mathsf{s})E^{(1)}(r_i)\Big). \end{aligned}$$

LHS of the above expression is the cost incurred to $\mathscr{F}_1$ when the strategy played is $(\pi_1(x), \pi_2^*(x))$ (refer to (46)).

Substituting for $D^{(1)}$, $E^{(1)}(r_i)$ and $C^{(1)}_{\pi_1^*,\pi_2^*}$, (from (11), (16) and (18), respectively) in the above expression and then rearranging, we can write

$$J^{(1)}_{\pi_1^*,\pi_2^*}(x) \quad \leq \quad \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[g_1(X_1,(A_{1,1},A_{2,1}))\Big] + \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[J^{(1)}_{\pi_1^*,\pi_2^*}(X_2)\Big]$$

Observe that, $J^{(1)}_{\pi_1^*,\pi_2^*}(\cdot)$ appears on the RHS of the above expression. Hence, inductively applying the above inequality K times, we obtain

$$J^{(1)}_{\pi_1^*,\pi_2^*}(x) \quad \leq \quad \sum_{k=1}^{K} \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[g_1(X_k,(A_{1,k},A_{2,k}))\Big] + \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[J^{(1)}_{\pi_1^*,\pi_2^*}(X_{K+1})\Big]$$

Taking limit as $K \to \infty$ in the above expression we obtain,

$$J^{(1)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(1)}_{\pi_1,\pi_2^*}(x) + \lim_{K\to\infty} \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[J^{(1)}_{\pi_1^*,\pi_2^*}(X_{K+1})\Big]. \tag{47}$$

Now, let $\mathcal{A}_{\boldsymbol{t}} = \{(\boldsymbol{t},\boldsymbol{t})\} \cup \{(\boldsymbol{t}, r_j) : j \in [n]\}$. $\mathcal{A}_{\boldsymbol{t}}$ is the set of all states, which are entered once $\mathscr{F}_1$ terminates. We will assume that the policy pair (π_1, π_2^*) is such that $\mathscr{F}_1$ will eventually terminate starting from any state x, i.e., $\lim_{K\to\infty} \mathbb{P}^x_{\pi_1,\pi_2^*}(X_K \in \mathcal{A}_{\boldsymbol{t}}) = 1$, or equivalently, for any $x' \notin \mathcal{A}_{\boldsymbol{t}}$, $\lim_{K\to\infty} \mathbb{P}^x_{\pi_1,\pi_2^*}(X_K = x') = 0$ (otherwise, with positive probability $\mathscr{F}_1$ will continue forever incurring a delay cost of τ at every stage yielding $J^{(1)}_{\pi_1^*,\pi_2^*}(x) = \infty$, so that the inequality $J^{(1)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(1)}_{\pi_1,\pi_2^*}(x)$ trivially holds). Using this along with $J^{(1)}_{\pi_1^*,\pi_2^*}(x^o) = 0$ for any $x^o \in \mathcal{A}_{\boldsymbol{t}}$, we can write

$$\begin{aligned} \lim_{K\to\infty} \mathbb{E}^x_{\pi_1,\pi_2^*}\Big[J^{(1)}_{\pi_1^*,\pi_2^*}(X_{K+1})\Big] \quad &= \quad \lim_{K\to\infty} \Bigg(\sum_{x^o \in \mathcal{A}_{\boldsymbol{t}}} \mathbb{P}^x_{\pi_1,\pi_2^*}(X_{K+1} = x^o) J^{(1)}_{\pi_1^*,\pi_2^*}(x^o) + \\ &\qquad\qquad \sum_{x' \notin \mathcal{A}_{\boldsymbol{t}}} \mathbb{P}^x_{\pi_1,\pi_2^*}(X_{K+1} = x') J^{(1)}_{\pi_1^*,\pi_2^*}(x') \Bigg) \\ &\overset{*}{=} \quad \sum_{x' \notin \mathcal{A}_{\boldsymbol{t}}} \lim_{K\to\infty} \Big(\mathbb{P}^x_{\pi_1,\pi_2^*}(X_{K+1} = x') J^{(1)}_{\pi_1^*,\pi_2^*}(x')\Big) \\ &\overset{o}{=} \quad \sum_{x' \notin \mathcal{A}_{\boldsymbol{t}}} \Big(\lim_{K\to\infty} \mathbb{P}^x_{\pi_1,\pi_2^*}(X_{K+1} = x')\Big)\; J^{(1)}_{\pi_1^*,\pi_2^*}(x') \\ &= \quad 0. \end{aligned}$$

Note that, in $*$ interchanging the limit and summation was possible because we have a finite sum (since our state space is finite). Also, since we have restricted ourselves to the class of stationary policies (recall Definition 1), $J^{(1)}_{\pi_1^*,\pi_2^*}(x')$ is not a function of the stage index K, which enables us to proceed to o. Finally, using the above in (47) we obtain, $J^{(1)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(1)}_{\pi_1,\pi_2^*}(x)$.

Similarly, for $\mathscr{F}_2$ it can be shown that $J^{(2)}_{\pi_1^*,\pi_2^*}(x) \leq J^{(2)}_{\pi_1^*,\pi_2}(x)$ for any π_2 and $x = (r_i, r_j)$. ■

Appendix C
Proof of Lemma 2

Lemma 2 will be an immediate consequence of the following result.

Lemma 7: Given an NEPP (π_1^*, π_2^*), for any $(r_i, r_j) \in \mathcal{X}$ we have,

$$J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, \boldsymbol{t}) \quad \leq \quad J^{(1)}_{\pi_1^*,\pi_2^*}(r_i, r_j), \tag{48}$$

$$J^{(2)}_{\pi_1^*,\pi_2^*}(\boldsymbol{t}, r_j) \quad \leq \quad J^{(2)}_{\pi_1^*,\pi_2^*}(r_i, r_j). \tag{49}$$

Proof: We will prove only (48); the proof of (49) is along similar lines. Since (π_1^*, π_2^*) is an NEPP, it follows that the policy π_1^* is the best response for $\mathscr{F}_1$ against the policy π_2^* of $\mathscr{F}_2$, i.e., for any $x \in \mathcal{X}$, $J^{(1)}_{\pi_1^*,\pi_2^*}(x) = \inf_{\pi_1} J^{(1)}_{\pi_1,\pi_2^*}(x)$. Thus $J^{(1)}_{\pi_1^*,\pi_2^*}(x)$ can be regarded as the optimal cost of the MDP problem, $MDP_1(\pi_2^*)$, obtained by fixing the policy π_2^* of $\mathscr{F}_2$.

For simplicity of notation we will denote $J^{(1)}_{\pi_1^*,\pi_2^*}(x)$ as $H^*(x)$. Thus for the states of the form $(r_i,\boldsymbol{t})$, $H^*(r_i,\boldsymbol{t})$ satisfies the following Bellman equation (this expression is same as the one in (10))

$$H^*(r_i,\boldsymbol{t}) = \min\Big\{-\eta_1 r_i, C_{\mathsf{c}}(r_i,\boldsymbol{t})\Big\}, \tag{50}$$

where

$$C_{\mathsf{c}}(r_i,\boldsymbol{t}) = \tau + \sum_{i'} p^{(1)}_{i'} H^*(r_{i'},\boldsymbol{t}) \tag{51}$$

is the expected cost of continuing, and $-\eta_1 r_i$ is the cost of stopping.

However, for states of the form (r_i,r_j) (where $\mathscr{F}_2$ is also competing for a relay), the optimality equation is more involved since the actions of $\mathscr{F}_2$ will now affect both costs (stopping and continuing) of $\mathscr{F}_1$. Defining $\epsilon = \pi_2^*(r_i,r_j,\mathsf{s})$ (ϵ is the probability with which $\mathscr{F}_2$ will stop when the state is (r_i,r_j)), Bellman equation for states of the form (r_i,r_j) can be written as

$$H^*(r_i,r_j) = \min\Big\{C_{\mathsf{s}}(r_i,r_j), C_{\mathsf{c}}(r_i,r_j)\Big\}, \tag{52}$$

where $C_{\mathsf{s}}(r_i,r_j)$ is the expected cost incurred by $\mathscr{F}_1$ for stopping when the state is (r_i,r_j), and $C_{\mathsf{c}}(r_i,r_j)$ is the expected cost of continuing.

The expression for $C_{\mathsf{s}}(r_i,r_j)$ is (recall that ν_ρ, $\rho=1,2$, is the probability that $\mathscr{F}_\rho$ gets the relay if both forwarders simultaneously choose to stop),

$$C_{\mathsf{s}}(r_i,r_j) \;=\; \epsilon\Big(\nu_1(-\eta_1 r_i) + \nu_2 C_{\mathsf{c}}(r_i,\boldsymbol{t})\Big) + (1-\epsilon)\Big(-\eta_1 r_i\Big). \tag{53}$$

The first term in the RHS of the above expression is the expected stopping cost incurred by $\mathscr{F}_1$ conditioned on the event that $\mathscr{F}_2$ also decides to stop. This can be understood as follows: suppose $\mathscr{F}_2$ also decides to stop (probability of which is ϵ), then w.p. ν_1, $\mathscr{F}_1$ gets the relay incurring a termination cost of $-\eta_1 r_i$, otherwise $\mathscr{F}_2$ gets the relay in which case $\mathscr{F}_1$ has to continue alone, the expected cost of which is $C_{\mathsf{c}}(r_i,\boldsymbol{t})$. The remaining term, $(1-\epsilon)(-\eta_1 r_i)$, in (53) is the stopping cost incurred to $\mathscr{F}_1$ when the action of $\mathscr{F}_2$ is to continue (which happens with probability $(1-\epsilon)$).

Similarly, the cost incurred by $\mathscr{F}_1$ for continuing, $C_{\mathsf{c}}(r_i,r_j)$, can be written as,

$$C_{\mathsf{c}}(r_i,r_j) \;=\; \epsilon\Big(\tau + \sum_{i'} p^{(1)}_{i'} H^*(r_{i'},\boldsymbol{t})\Big) + (1-\epsilon)\Big(\tau + \sum_{i',j'} p_{i',j'} H^*(r_{i'},r_{j'})\Big). \tag{54}$$

Now, returning to (50) and (52), H^* can be expressed as the fixed point of a mapping T which is, for a function $H(\cdot,\cdot)$, given by,

$$\begin{aligned} TH(r_i,\boldsymbol{t}) &= \min\Big\{-\eta_1 r_i, C^H_{\mathsf{c}}(r_i,\boldsymbol{t})\Big\} \\ TH(r_1,r_2) &= \min\Big\{C^H_{\mathsf{s}}(r_i,r_j), C^H_{\mathsf{c}}(r_i,r_j)\Big\}, \end{aligned}$$

where the expressions for $C^H_{\mathsf{c}}(r_i,\boldsymbol{t})$, $C^H_{\mathsf{s}}(r_i,r_j)$ and $C^H_{\mathsf{c}}(r_i,r_j)$ is similar to that of $C_{\mathsf{c}}(r_i,\boldsymbol{t})$, $C_{\mathsf{s}}(r_i,r_j)$ and $C_{\mathsf{c}}(r_i,r_j)$ (in (51) (53) and (54), respectively) with H^* replaced by the given function H. Inductively define $H_k = TH_{k-1}$ with $H_0 \equiv 0$ (i.e., $H_0(x)=0$ for all $x\in\mathcal{X}$). Since $MDP_1(\pi_2^*)$ is an optimal stopping problem [39] it follows that $H_k \to H^*$ (this is the value iteration algorithm). Hence, to complete the proof we will show that $H_k(r_i,\boldsymbol{t}) \le H_k(r_i,r_j)$ whenever $H_{k-1}(r_i,\boldsymbol{t}) \le H_{k-1}(r_i,r_j)$.

Suppose, for some $k\ge 1$, $H_{k-1}(r_i,\boldsymbol{t}) \le H_{k-1}(r_i,r_j)$ for all $(r_i,r_j)\in\mathcal{X}$ (this holds trivially for $k=1$). First consider the case where, $-\eta_1 r_i \le C^{H_{k-1}}_{\mathsf{c}}(r_i,\boldsymbol{t})$ (i.e., it is optimal to stop when the state is $(r_i,\boldsymbol{t})$).

- Then from (53) we obtain $-\eta_1 r_i \le C^{H_{k-1}}_{\mathsf{s}}(r_i,r_j)$.
- Also, from the induction hypothesis we have

$$\begin{aligned} \sum_{i'} p^{(1)}_{i'} H_{k-1}(r_{i'},\boldsymbol{t}) &= \sum_{i',j'} p_{i',j'} H_{k-1}(r_{i'},\boldsymbol{t}) \\ &\le \sum_{i',j'} p_{i',j'} H_{k-1}(r_{i'},r_{j'}). \end{aligned}$$

  Using the above in (54) and recalling (51) we can write

$$\begin{aligned} C^{H_{k-1}}_{\mathsf{c}}(r_i,r_j) &\ge \tau + \sum_{i'} p^{(1)}_{i'} H^*(r_{i'},\boldsymbol{t}) \\ &= C_{\mathsf{c}}(r_i,\boldsymbol{t}) \\ &\ge -\eta_1 r_i. \end{aligned}$$

Thus we have,

$$\begin{aligned} H_k(r_i,\boldsymbol{t}) &= \min\Big\{-\eta_1 r_i, C_{\mathsf{c}}^{H_{k-1}}(r_i,\boldsymbol{t})\Big\} \\ &= -\eta_1 r_i \\ &\leq \min\Big\{C_{\mathsf{s}}^{H_{k-1}}(r_i,r_j), C_{\mathsf{c}}^{H_{k-1}}(r_i,r_j)\Big\} \\ &= H_k(r_i,r_j). \end{aligned}$$

Similarly for the other case, i.e., when $-\eta_1 r_i > C_{\mathsf{c}}^{H}(r_i,\boldsymbol{t})$, we can show that both the costs, $C_{\mathsf{s}}^{H_{k-1}}(r_i,r_j)$ and $C_{\mathsf{s}}^{H_{k-1}}(r_i,r_j)$, are less than $C_{\mathsf{c}}^{H}(r_i,\boldsymbol{t})$ again yielding $H_k(r_i,\boldsymbol{t}) \leq H_k(r_i,r_j)$. ∎

We are now ready to prove Lemma 2.

Lemma 2: For an NEPP, (π_1^*,π_2^*), the various costs are ordered as follows:

$$D^{(1)} \leq C^{(1)}_{\pi_1^*,\pi_2^*} \text{ and } D^{(2)} \leq C^{(2)}_{\pi_1^*,\pi_2^*}.$$

Proof: Recalling the cost expressions of $D^{(1)}$ and $C^{(1)}_{\pi_1^*,\pi_2^*}$ (from (11) and (18), respectively) we can write,

$$\begin{aligned} D^{(1)} &= \tau + \sum_{i'} p_{i'}^{(1)} J^{(1)}_{\pi_1^*,\pi_2^*}(r_{i'},\boldsymbol{t}) \\ &= \tau + \sum_{i',j'} p_{i',j'} J^{(1)}_{\pi_1^*,\pi_2^*}(r_{i'},\boldsymbol{t}) \\ &\overset{*}{\leq} \tau + \sum_{i',j'} p_{i',j'} J^{(1)}_{\pi_1^*,\pi_2^*}(r_{i'},r_{j'}) \\ &= C^{(1)}_{\pi_1^*,\pi_2^*}, \end{aligned}$$

where $*$ is due to Lemma 7. Similarly, one can show that $D^{(2)} \leq C^{(2)}_{\pi_1^*,\pi_2^*}$. ∎

Appendix D
Obtaining NE Strategies for the Static Game in Table 4

For convenience, let us first recall the game in Table 4.

	c	s
c	$C^{(1)}_{\pi_1^*,\pi_2^*}, C^{(2)}_{\pi_1^*,\pi_2^*}$	$D^{(1)}, -\eta_2 r_j$
s	$-\eta_1 r_i, D^{(2)}$	$E^{(1)}(r_i), E^{(2)}(r_j)$

TABLE 6
Static stage game.

Since only two actions (namely s and c) are available to each forwarder, a strategy used by $\mathscr{F}_1$ can be conveniently represented by $\sigma_1 \in [0,1]$, where σ_1 is the probability that $\mathscr{F}_1$ will choose action s. Similarly, $\sigma_2 \in [0,1]$ is the probability that $\mathscr{F}_2$ will choose action s. Given a strategy pair (σ_1,σ_2) the expected cost (obtained from Table 4) incurred by $\mathscr{F}_1$ can be expressed as

$$U_1(\sigma_1,\sigma_2) = \sigma_1 A_{\sigma_2} + B_{\sigma_2}, \tag{55}$$

where

$$A_{\sigma_2} = (1-\sigma_2)\Big(-\eta_1 r_i - C^{(1)}\Big) + \sigma_2\Big(E^{(1)}(r_i) - D^{(1)}\Big) \tag{56}$$

and

$$B_{\sigma_2} = (1-\sigma_2)C^{(1)} + \sigma_2 D^{(1)}.$$

Let $\sigma_1^*(\sigma_2)$ denote the set of all best responses of $\mathscr{F}_1$ to the strategy σ_2 of $\mathscr{F}_2$, i.e.,

$$\sigma_1^*(\sigma_2) = \arg\min_{\sigma_1 \in [0,1]} U_1(\sigma_1,\sigma_2). \tag{57}$$

Since $U_1(\sigma_1,\sigma_2)$ is linear in σ_1 it follows that, $\sigma_1^*(\sigma_2) = \{0\}$ whenever $A_{\sigma_2} > 0$, $\sigma_1^*(\sigma_2) = \{1\}$ whenever $A_{\sigma_2} < 0$, and $\sigma_1^*(\sigma_2) = [0,1]$ whenever $A_{\sigma_2} = 0$. We make use of these observations in the proof of our next lemma. First for convenience let us denote the thresholds $\frac{C^{(1)}}{-\eta_1}$ and $\frac{C^{(2)}}{-\eta_2}$ by $\zeta^{(1)}$ and $\zeta^{(2)}$, respectively. Recall that we already have, $\alpha^{(1)} = \frac{D^{(1)}}{-\eta_1}$ and $\alpha^{(2)} = \frac{D^{(2)}}{-\eta_2}$. The inequalities in (20) would (since, $-\eta < 0$) imply that $\zeta^{(1)} \leq \alpha^{(1)}$ and $\zeta^{(2)} \leq \alpha^{(2)}$.

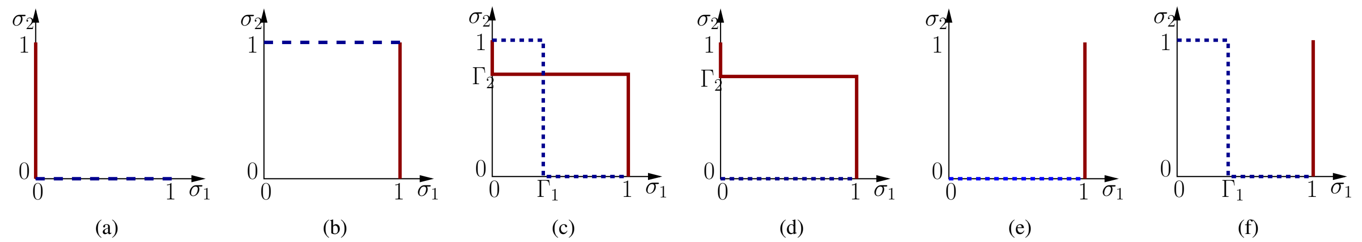


Fig. 5. Plot of best response curves, $\sigma_1^*(\sigma_2)$ and $\sigma_2^*(\sigma_1)$, for (r_i, r_j) in different regions. In each of these figures, the solid red curve is $\sigma_1^*(\sigma_2)$ and the dashed blue curve is $\sigma_2^*(\sigma_1)$. (a) $(r_i, r_j) \in \mathcal{R}_1$, (b) $(r_i, r_j) \in \mathcal{R}_5$, (c) $(r_i, r_j) \in \mathcal{R}_4$, (d) $(r_i, r_j) \in \mathcal{R}_{2a}$, (e) $(r_i, r_j) \in \mathcal{R}_{2b}$, and (f) $(r_i, r_j) \in \mathcal{R}_{2c}$.

Lemma 8: Suppose $\nu_1 \in (0,1)$ and $D^{(1)} < C^{(1)}$, then

1) If $r_i < \zeta^{(1)}$ then $\sigma_1^*(\sigma_2) = \{0\}$ for all $\sigma_2 \in [0,1]$.
2) If $r_i > \alpha^{(1)}$ then $\sigma_1^*(\sigma_2) = \{1\}$ for all $\sigma_2 \in [0,1]$.
3) If $\zeta^{(1)} \leq r_i \leq \alpha^{(1)}$ then defining

$$\Gamma_2 \quad = \quad \frac{-\eta_1 r_i - C^{(1)}}{\Big(-\eta_1 r_i - C^{(1)}\Big) - \Big(E^{(1)}(r_i) - D^{(1)}\Big)} \tag{58}$$

we have: (i) $\sigma_1^*(\sigma_2) = \{1\}$ for $\sigma_2 < \Gamma_2$, (ii) $\sigma_1^*(\sigma_2) = \{0\}$ for $\sigma_2 > \Gamma_2$, and (iii) $\sigma_1^*(\Gamma_2) = [0,1]$.

Proof of Part 1: We will show that $A_{\sigma_2} > 0$ for any $\sigma_2 \in [0,1]$. Then the proof follows immediately since, $U_1(\sigma_1, \sigma_2) = \sigma_1 A_{\sigma_2} + B_{\sigma_2}$, is linear in σ_1.

Let us recall the expression for A_{σ_2},

$$A_{\sigma_2} = (1-\sigma_2)\Big(-\eta_1 r_i - C^{(1)}\Big) + \sigma_2\Big(C^{(1)}(r_i) - D^{(1)}\Big), \tag{59}$$

where $E^{(1)}(r_i) = \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}$ (see (16)). It is already given that $r_i < \frac{C^{(1)}}{-\eta_1}$, or

$$\Big(-\eta_1 r_i - C^{(1)}\Big) > 0. \tag{60}$$

Since $D^{(1)} < C^{(1)}$ we also have $-\eta_1 r_i > D^{(1)}$ which gives $E^{(1)}(r_i) > D^{(1)}$ (this is where $\nu_1 \in (0,1)$ is required), i.e, $\Big(E_{(1)}(r_i) - D^1\Big) > 0$. Using this along with inequality (60) we obtain the desired result.

(Proof of Part 2) Similar to the previous part, the proof follows once we show that $A_{\sigma_2} < 0$ for all $\sigma_2 \in [0,1]$. Since $r_i > \frac{D^{(1)}}{-\eta_1}$ and $D^{(1)} < C^{(1)}$, we obtain $\Big(-\eta_1 r_i - C^{(1)}\Big) < 0$ and $\Big(E^{(1)}(r_i) - D^{(1)}\Big) < 0$. Using these in (59) we obtain $A_{\sigma_2} < 0$.

(Proof of Part 3) Again, since $U_1(\sigma_1, \sigma_2)$ is linear in σ_1, we have to show that $A_{\sigma_2} < 0$ whenever $\sigma_2 < \Gamma_2$, $A_{\sigma_2} > 0$ whenever $\sigma_2 > \Gamma_2$, and $A_{\Gamma_2} = 0$.

Suppose $\sigma_2 \in [0,1]$ is such that $\sigma_2 < \Gamma_2$ (thus $\Gamma_2 \in (0,1]$), then recalling the expression for Γ_2 we can write,

$$\sigma_2 < \frac{-\eta_1 r_i - C^{(1)}}{\Big(-\eta_1 r_i - C^{(1)}\Big) - \Big(E^{(1)}(r_i) - D^{(1)}\Big)}. \tag{61}$$

It is important to note that, since $\frac{C^{(1)}}{-\eta_1} \leq r_i \leq \frac{D^{(1)}}{-\eta_1}$ with $D^{(1)} < C^{(1)}$ and $\nu_1 \in (0,1)$, the denominator in the RHS of the above expression is strictly negative. Thus, rearranging (61) we obtain $A_{\sigma_2} < 0$ so that $\sigma_1^*(\sigma_2) = \{1\}$.

Similarly, when $\sigma_2 \in [0,1]$ is such that $\sigma_2 > \Gamma_2$ (in which case $\Gamma_2 \in [0,1)$), then reversing the inequality in (61) and rearranging we obtain $A_{\sigma_2} > 0$ so that $\sigma_1^*(\sigma_2) = \{0\}$.

Finally, substituting for Γ_2 in the expression for A_{Γ_2} will yield $A_{\Gamma_2} = 0$ implying that any $\sigma_1 \in [0,1]$ is a best response against the strategy Γ_2 played by $\mathscr{F}_2$. Hence $\sigma_1^*(\Gamma_2) = [0,1]$. ■

Remark: The condition imposed on ν_1 and $C^{(1)}$ in the above lemma is only to avoid the less interesting boundary cases. Also, note that Γ_2 is a function of the reward r_i to $\mathscr{F}_1$. For notational simplicity we do not show r_i as an argument of Γ_2.

Similarly, for $\mathscr{F}_2$ we can define $\sigma_2^*(\sigma_1)$ as the set of all best responses against the strategy σ_1 played by $\mathscr{F}_1$, and obtain a result analogous to that in Lemma 8, but with quantities corresponding to $\mathscr{F}_1$ replaced by that corresponding to $\mathscr{F}_2$, e.g., for instance, $\zeta^{(1)}$ replaced by $\zeta^{(2)}$, $\alpha^{(1)}$ by $\alpha^{(2)}$, Γ_2 by Γ_1 where

$$\Gamma_1 = \frac{-\eta_2 r_j - C^{(2)}}{\Big(-\eta_2 r_j - C^{(2)}\Big) - \Big(E^{(2)}(r_j) - D^{(2)}\Big)}. \tag{62}$$

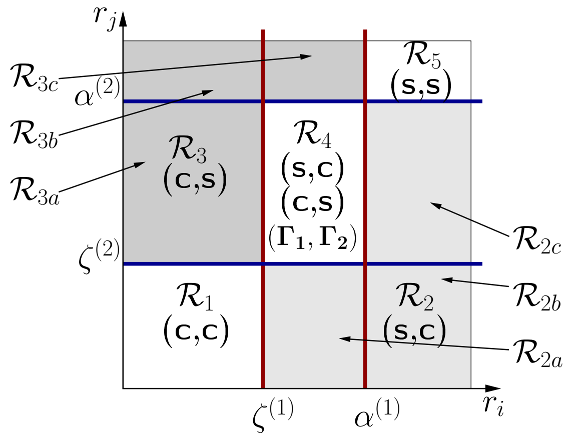


Fig. 6. Illustration of the various regions along with the NE strategy corresponding to these regions.

Now, for any (r_i, r_j) the points of intersection between the best response curves $\sigma_1^*(\sigma_2)$ and $\sigma_2^*(\sigma_1)$ constitutes the NE strategies of the game in Table 4. For instance, as shown in Fig. 5(a), when (r_i, r_j) is such that $r_i < \zeta^{(1)}$ and $r_j < \zeta^{(2)}$ (i.e., $(r_i, r_j) \in \mathcal{R}_1$; see Fig. 6) then the only point of intersection is $(0, 0)$ so that (c, c) is the only NE strategy in this region. Similarly when $(r_i, r_j) \in \mathcal{R}_5$ then (s, s) is the only NE strategy (see Fig. 5(b)). An interesting case is when $(r_i, r_j) \in \mathcal{R}_4$ (see Fig. 5(c)) where there are multiple NE strategies, namely, (s, c), (c, s) and the mixed strategy (Γ_1, Γ_2) (which depends on the reward pair (r_i, r_j); see remark following Lemma 8).

The region $\mathcal{R}_2$ is written as a union of three disjoint regions, $\mathcal{R}_{2a}$, $\mathcal{R}_{2b}$ and $\mathcal{R}_{2c}$. However, as shown in Fig. 5(d) to 5(f), the best response curves for (r_i, r_j) in each of these sub-regions intersect at $(1, 0)$. Hence (s, c) is the NE strategy in the union region $\mathcal{R}_2$. Similarly, (c, s) is the NE strategy in region $\mathcal{R}_3$ which is also composed of three sub-regions.

$\mathcal{R}_1 = \left\{(r_i, r_j) : r_i < \zeta^{(1)}, r_j < \zeta^{(2)}\right\}$
$\mathcal{R}_2 = \mathcal{R}_{2a} \cup \mathcal{R}_{2b} \cup \mathcal{R}_{2c}$ where $\mathcal{R}_{2a} = \left\{(r_i, r_j) : \zeta^{(1)} \le r_i \le \alpha^{(1)}, r_j < \zeta^{(2)}\right\}$ $\mathcal{R}_{2b} = \left\{(r_i, r_j) : r_i > \alpha^{(1)}, r_j < \zeta^{(2)}\right\}$ $\mathcal{R}_{2c} = \left\{(r_i, r_j) : r_i > \alpha^{(1)}, \zeta^{(2)} \le r_j \le \alpha^{(2)}\right\}$
$\mathcal{R}_3 = \mathcal{R}_{3a} \cup \mathcal{R}_{3b} \cup \mathcal{R}_{3c}$ where $\mathcal{R}_{3a} = \left\{(r_i, r_j) : r_i < \zeta^{(1)}, \zeta^{(2)} \le r_j \le \alpha^{(2)}\right\}$ $\mathcal{R}_{3b} = \left\{(r_i, r_j) : r_i < \zeta^{(1)}, r_j > \alpha^{(2)}\right\}$ $\mathcal{R}_{3c} = \left\{(r_i, r_j) : \zeta^{(1)} \le r_i \le \alpha^{(1)}, r_j > \alpha^{(2)}\right\}$
$\mathcal{R}_4 = \left\{(r_i, r_j) : \zeta^{(1)} \le r_i \le \alpha^{(1)}, \zeta^{(2)} \le r_j \le \alpha^{(2)}\right\}$
$\mathcal{R}_5 = \left\{(r_i, r_j) : r_i > \alpha^{(1)}, r_j > \alpha^{(2)}\right\}$

TABLE 7
FORMAL DEFINITION OF VARIOUS REGIONS DEPICTED IN FIG. 6.

We have thus identified a partition of the set $\{(r_i, r_j) : i, j \in [n]\}$ into five regions such that the set of NE strategies corresponding to each region are different. These regions along with the corresponding NE strategies are depicted in Fig. 6. A formal definition of the various regions is available in Table 7. Note that, these regions depend on the cost pair $\mathbf{C} = (C^{(1)}, C^{(2)})$; for simplicity we have not shown this explicitly in Fig. 6 and in Table 7.

APPENDIX E
PROOF OF THEOREM 3

Theorem 3: Given a PO policy pair $(\overline{\pi}_1^*, \overline{\pi}_2^*)$, construct a strategy vector pair $\{(f_\ell^*, g_\ell^*) : \ell \in \mathcal{L}\}$ as follows: $f_\ell^*(r_i) = \overline{\pi}_1^*(r_i, \ell)$ and $g_\ell^*(r_j) = \overline{\pi}_2^*(\ell, r_j)$ for all $i, j \in [n]$. Now, suppose for each ℓ, (f_ℓ^*, g_ℓ^*) is a NE vector for $\mathcal{G}(\overline{\pi}_1^*, \overline{\pi}_2^*)$ such that,

$$\begin{aligned} \min\left\{C^{(1)}_{\mathsf{s}, g_\ell^*}(r_i), C^{(1)}_{\mathsf{c}, g_\ell^*}\right\} &= G^{(1)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(r_i, \ell), \text{ and} \\ \min\left\{C^{(2)}_{\mathsf{s}, f_\ell^*}(r_j), C^{(2)}_{\mathsf{c}, f_\ell^*}\right\} &= G^{(2)}_{\overline{\pi}_1^*, \overline{\pi}_2^*}(\ell, r_j). \end{aligned}$$

Then $(\overline{\pi}_1^*, \overline{\pi}_2^*)$ is a PO-NEPP.

Proof: Given the policy pair $(\overline{\pi}_1^*, \overline{\pi}_2^*)$ as in the hypothesis, let $\overline{\pi}_1$ be any PO policy. We will show that $G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) \leq G^{(1)}_{\overline{\pi}_1,\overline{\pi}_2^*}(r_i,\ell)$; the proof of, $G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) \leq G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2}(r_i,\ell)$ for any $\overline{\pi}_2$, is along similar lines.

Since $f_\ell^* = BR_1(g_\ell^*)$, for the Bayesian game $\mathcal{G}(\overline{\pi}_1^*, \overline{\pi}_2^*)$, the expected cost incurred to $\mathscr{F}_1$ when its observation is (r_i,ℓ) is $\min\left\{C^{(1)}_{\mathsf{s},g_\ell^*}(r_i), C^{(1)}_{\mathsf{c},g_\ell^*}\right\}$. Hence, using (37) we can write

$$\begin{aligned} G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) &\leq C^{(1)}_{\overline{\pi}_1(r_i,\ell),g_\ell^*}(r_i) \\ &= \mathbb{E}^{(r_i,\ell)}_{\overline{\pi}_1,\overline{\pi}_2^*}\Big[g_1(X_1,(A_{1,1},A_{2,1}))\Big] + \mathbb{E}^{(r_i,\ell)}_{\overline{\pi}_1,\overline{\pi}_2^*}\Big[G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(O_{1,2})\Big]. \end{aligned}$$

Applying the above inequality K times we obtain

$$G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) \leq \sum_{k=1}^{K}\mathbb{E}^{(r_i,\ell)}_{\overline{\pi}_1,\overline{\pi}_2^*}\Big[g_1(X_k,(A_{1,k},A_{2,k}))\Big] + \mathbb{E}^{(r_i,\ell)}_{\overline{\pi}_1,\overline{\pi}_2^*}\Big[G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(O_{1,K+1})\Big]. \tag{63}$$

Again, as in the proof of Theorem 1 Part-(b), we will assume that the PO policy pair $(\overline{\pi}_1, \overline{\pi}_2^*)$ is such that using this policy pair $\mathscr{F}_1$ will eventually terminate starting from any observation o_1, i.e.,

$$\lim_{K\to\infty}\mathbb{P}^{o_1}_{\overline{\pi}_1,\overline{\pi}_2^*}(O_{1,K} = \boldsymbol{t}) = 1. \tag{64}$$

Hence we have

$$\begin{aligned} \lim_{K\to\infty}\mathbb{E}^{(r_i,\ell)}_{\overline{\pi}_1,\overline{\pi}_2^*}\Big[G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(O_{1,K+1})\Big] &= G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(\boldsymbol{t}) \\ &= 0. \end{aligned}$$

Using the above and recalling (28) while taking $\lim_{K\to\infty}$ in (63) we obtain $G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) \leq G^{(1)}_{\overline{\pi}_1,\overline{\pi}_2^*}(r_i,\ell)$.

Finally, suppose the PO policy pair $(\overline{\pi}_1, \overline{\pi}_2^*)$ does not satisfy (64), then there is a positive probability that $\mathscr{F}_1$ will continue forever yielding $G^{(1)}_{\overline{\pi}_1,\overline{\pi}_2^*}(r_i,\ell) = \infty$. Thus for this case, $G^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}(r_i,\ell) \leq G^{(1)}_{\overline{\pi}_1,\overline{\pi}_2^*}(r_i,\ell)$, trivially holds. ■

APPENDIX F
PROOF OF LEMMA 5

Lemma 5: (1) Let $\Psi_\ell, \Psi_\ell^o \in \mathcal{A}_0$ be two thresholds of $\mathscr{F}_2$ such that $\Psi_\ell < \Psi_\ell^o$, then the best response of $\mathscr{F}_1$ to these are ordered as, $BR_1(\Psi_\ell) \geq BR_1(\Psi_\ell^o)$. (2) Similarly, if $\Phi_\ell, \Phi_\ell^o \in \mathcal{A}_0$ are two thresholds of $\mathscr{F}_1$ such that $\Phi_\ell < \Phi_\ell^o$ then $BR_2(\Phi_\ell) \geq BR_2(\Phi_\ell^o)$.

Proof: For convenience, first let us recall the expressions of the costs $C^{(1)}_{\mathsf{s},g_\ell}(r_i)$ and $C^{(1)}_{\mathsf{c},g_\ell}$ from (35) and (36) (since the given PO policy pair is $(\overline{\pi}_1^*, \overline{\pi}_2^*)$, these costs correspond to the Bayesian game $\mathcal{G}(\overline{\pi}_1^*, \overline{\pi}_2^*)$):

$$C^{(1)}_{\mathsf{s},g_\ell}(r_i) = \widetilde{g}_\ell(-\eta_1 r_i) + (1-\widetilde{g}_\ell)E^{(1)}(r_i) \tag{65}$$

$$C^{(1)}_{\mathsf{c},g_\ell} = \widetilde{g}_\ell\overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*} + (1-\widetilde{g}_\ell)D^{(1)}. \tag{66}$$

Also, recall from (33) that the cost of continuing alone is less than the cost of continuing along with the competing forwarder, i.e.,

$$D^{(1)} \leq \overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*} \tag{67}$$

We will only prove Part-(1); the proof of Part-(2) is similar. Let g_ℓ and g_ℓ^o be the threshold vectors of $\mathscr{F}_2$ whose corresponding thresholds are Ψ_ℓ and Ψ_ℓ^o, respectively. Given that $\Psi_\ell < \Psi_\ell^o$, to prove $BR_1(\Psi_\ell) \geq BR_1(\Psi_\ell^o)$ it is sufficient to show that, for any r_i, $C^{(1)}_{\mathsf{s},g_\ell}(r_i) \leq C^{(1)}_{\mathsf{c},g_\ell}$ implies $C^{(1)}_{\mathsf{s},g_\ell^o}(r_i) \leq C^{(1)}_{\mathsf{c},g_\ell^o}$.

Let us begin with an r_i such that $C^{(1)}_{\mathsf{s},g_\ell}(r_i) \leq C^{(1)}_{\mathsf{c},g_\ell}$, or alternatively, (recall (65) and (66)) r_i is such that,

$$\widetilde{g}_\ell(-\eta_1 r_i) + (1-\widetilde{g}_\ell)E^{(1)}(r_i) \leq \widetilde{g}_\ell\overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*} + (1-\widetilde{g}_\ell)D^{(1)}.$$

Substituting $E^{(1)}(r_i) = \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)}$ in the above expression, and then simplifying we obtain,

$$-\eta_1 r_i \leq \frac{\widetilde{g}_\ell\overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*} + (1-\widetilde{g}_\ell)\nu_1 D^{(1)}}{\widetilde{g}_\ell + (1-\widetilde{g}_\ell)\nu_1}$$

Thus $-\eta_1 r_i$, being less than the convex combination of the costs $\overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}$ and $D^{(1)}$, is less than both of these. Further since $D^{(1)} \leq \overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}$, there are only two cases which are possible: $-\eta_1 r_i < D^{(1)} \leq \overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}$, and $D^{(1)} \leq -\eta_1 r_i \leq \overline{C}^{(1)}_{\overline{\pi}_1^*,\overline{\pi}_2^*}$. We will consider these two cases separately below.

Case-1: Suppose $-\eta_1 r_i < D^{(1)} \leq \overline{C}^{(1)}_{\pi_1^*,\pi_2^*}$, then

$$\begin{aligned} E^{(1)}(r_i) &= \nu_1(-\eta_1 r_i) + \nu_2 D^{(1)} \\ &\leq D^{(1)}. \end{aligned}$$

Using the above two inequalities in the expression of $C^{(1)}_{\mathsf{s},g^o_\ell}(r_i)$, and then comparing with $C^{(1)}_{\mathsf{c},g^o_\ell}$ we obtain $C^{(1)}_{\mathsf{s},g^o_\ell}(r_i) \leq C^{(1)}_{\mathsf{c},g^o_\ell}$.

Case-2: Suppose $D^{(1)} \leq -\eta_1 r_i \leq \overline{C}^{(1)}_{\pi_1^*,\pi_2^*}$. Then we have $E^{(1)}(r_i) \geq D^{(1)}$. Define $\kappa(p)$ for $p \in [0,1]$ as,

$$\kappa(p) = p\Big(- \eta_1 r_i - E^{(1)}(r_i) - \overline{C}^{(1)}_{\pi_1^*,\pi_2^*} + D^{(1)}\Big) + \Big(E^{(1)}(r_i) - D^{(1)}\Big). \quad (68)$$

Since $-\eta_1 r_i \leq \overline{C}^{(1)}_{\pi_1^*,\pi_2^*}$ and $E^{(1)}(r_i) \geq D^{(1)}$ we have, $\kappa(p)$ is decreasing in p. Hence we can write $\kappa(\widetilde{g}^o_\ell) \leq \kappa(\widetilde{g}_\ell)$ because, with $\Psi_\ell < \Psi^o_\ell$ we have,

$$\widetilde{g}_\ell = \sum_{j=1}^{\Psi_\ell} p^{(2)}_{j|\ell} \leq \sum_{j=1}^{\Psi^o_\ell} p^{(2)}_{j|\ell} = \widetilde{g}^o_\ell. \quad (69)$$

Finally, rearranging the terms in (68) one can obtain,

$$\begin{aligned} C^{(1)}_{\mathsf{s},g^o_\ell}(r_i) - C^{(1)}_{\mathsf{c},g^o_\ell} &= \kappa(\widetilde{g}^o_\ell) \\ &\leq \kappa(\widetilde{g}_\ell) \\ &= C^{(1)}_{\mathsf{s},g_\ell}(r_i) - C^{(1)}_{\mathsf{c},g_\ell} \\ &\overset{*}{\leq} 0, \end{aligned}$$

where $*$ is because we started with an r_i such that, $C^{(1)}_{\mathsf{s},g_\ell}(r_i) \leq C^{(1)}_{\mathsf{c},g_\ell}$. ∎

Appendix G
Obtaining Cooperative Optimal Policy Pair

We will first prove Lemma 6.

Lemma 6: The policy pair $(\pi_1^\gamma, \pi_2^\gamma)$ is *Pareto optimal*, i.e., for any other policy (π_1,π_2),

(1) if $C^{(1)}_{\pi_1,\pi_2} < C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}$ then $C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma} < C^{(2)}_{\pi_1,\pi_2}$, and
(2) if $C^{(2)}_{\pi_1,\pi_2} < C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma}$ then $C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma} < C^{(1)}_{\pi_1,\pi_2}$.

Proof: We will prove Part-(1); the proof of Part-(2) is similar. Since $(\pi_1^\gamma,\pi_2^\gamma)$ is optimal for the problem in (40) we can write, for any policy pair (π_1,π_2),

$$\gamma C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma} + (1-\gamma) C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma} \leq \gamma C^{(1)}_{\pi_1,\pi_2} + (1-\gamma) C^{(2)}_{\pi_1,\pi_2},$$

rewriting which we obtain

$$C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma} \leq \frac{\gamma}{1-\gamma}\Big(C^{(1)}_{\pi_1,\pi_2} - C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}\Big) + C^{(2)}_{\pi_1,\pi_2}.$$

Now, if $C^{(1)}_{\pi_1,\pi_2} < C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}$ then from the above expression we have $C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma} < C^{(2)}_{\pi_1,\pi_2}$. ∎

We now proceed to obtain $(C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}, C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma})$ by formulating the problem in (40) as an MDP. The state space, action space and the state transitions remain same as in Section V. However the one-step costs have to be appropriately modified to take into account the multiplier γ. Without writing down all the details we will proceed to the Bellman equation. The one-step costs will be evident from these.

For states of the form $(r_i,\boldsymbol{t})$

$$J^*(r_i,\boldsymbol{t}) = \min\Big\{ -\gamma\eta_1 r_i, \gamma\tau + \sum_{i'} p^{(1)}_{i'} J^*(r'_i,\boldsymbol{t})\Big\}. \quad (70)$$

The first term in the $\min$ expression above corresponds to the cost of the joint-action (s,s) or (s,c) (since $\mathscr{F}_2$ has terminated, its action is irrelavent), and the second term is the expected cost of choosing (c,s) (or (c,c)). Similarly, when the state is $(\boldsymbol{t},r_j)$ we can write

$$J^*(\boldsymbol{t},r_j) = \min\Big\{ -(1-\gamma)\eta_2 r_j, (1-\gamma)\tau + \sum_{j'} p^{(2)}_{j'} J^*(\boldsymbol{t},r'_j)\Big\}. \quad (71)$$

The more interesting case is when both forwarders are still competing, i.e., when the state is of the form (r_i, r_j), where the optimality equation is

$$J^*(r_i, r_j) = \min\Big\{C_{\mathsf{s,c}}, C_{\mathsf{c,s}}, C_{\mathsf{s,s}}, C_{\mathsf{c,c}}\Big\}; \tag{72}$$

C_{a_1,a_2} is the expected cost (one-step + future cost-to-go) of choosing the joint-action (a_1, a_2). When the joint-action chosen is (s, c), since $\mathscr{F}_1$ stops and $\mathscr{F}_2$ continues the one-step cost is $\Big(-\gamma\eta_1 r_i + (1-\gamma)\tau\Big)$. The subsequent state is of the form $(\boldsymbol{t}, r_{j'})$ w.p. $p^{(2)}_{j'}$. Hence the expression for $C_{\mathsf{s,c}}$ can be written as

$$C_{\mathsf{s,c}} = -\gamma\eta_1 r_i + (1-\gamma)\tau + \sum_{j'} p^{(2)}_{j'} J^*(\boldsymbol{t}, r_{j'}). \tag{73}$$

Similarly $C_{\mathsf{c,s}}$ can be written as

$$C_{\mathsf{c,s}} = \gamma\tau - (1-\gamma)\eta_2 r_j + \sum_{i'} p^{(1)}_{i'} J^*(r_{i'}, \boldsymbol{t}). \tag{74}$$

When both forwarders decide to stop then w.p. ν_1, $\mathscr{F}_1$ gets the relay in which case the cost incurred is $C_{\mathsf{s,c}}$; otherwise, w.p. ν_2, $\mathscr{F}_2$ get the relay and the cost incurred is $C_{\mathsf{c,s}}$. Hence

$$C_{\mathsf{s,s}} = \nu_1 C_{\mathsf{s,c}} + \nu_2 C_{\mathsf{c,s}}. \tag{75}$$

Finally, when both forwarders continue the one-step cost is simply $\Big(\gamma\tau + (1-\gamma)\tau\Big) = \tau$ and the subsequent state is still of the form $(r_{i'}, r_{j'})$. Thus we can write

$$C_{\mathsf{c,c}} = \tau + \sum_{i',j'} p_{i',j'} J^*(r_{i'}, r_{j'}). \tag{76}$$

From (75) note that $C_{\mathsf{s,s}} \geq \min\{C_{\mathsf{s,c}}, C_{\mathsf{c,s}}\}$ which means that the joint-action (s, s) can never be optimal. Thus, under cooperation the forwarders never compete for a relay; either $\mathscr{F}_1$ will choose the relay, or $\mathscr{F}_2$ will choose, or both continue. Expression (72) can therefore be simplified as

$$J^*(r_i, r_j) = \min\Big\{C_{\mathsf{s,c}}, C_{\mathsf{c,s}}, C_{\mathsf{c,c}}\Big\}. \tag{77}$$

One can perform value iteration to solve for J^* in (70), (71) and (77). Given J^* it is easy to obtain the optimal policy $(\pi_1^\gamma, \pi_2^\gamma)$ (simply choose the joint-action that minimizes the RHS of these expressions, breaking ties arbitrarily) and then the cost pair, $(C^{(1)}_{\pi_1^\gamma,\pi_2^\gamma}, C^{(2)}_{\pi_1^\gamma,\pi_2^\gamma})$.